

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: April 23, 2003, 12:33:40 ; Search time 9.65517 seconds  
(without alignments)  
79.654 Million cell updates/sec

Title: US-09-723-722A-81

Perfect score: 34

Sequence: 1 EVMXVAEF 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	85.3	195	2 C97255	thymidine kinase (
2	29	85.3	236	2 B70728	hypothetical prote
3	28	82.4	594	2 A10673	probable hydrolase
4	28	82.4	1256	2 AB2042	hypothetical prote
5	28	82.4	2638	1 A42545	genome polyprotein
6	27	79.4	181	2 T11902	NADH2 dehydrogenas
7	27	79.4	184	2 C69133	DNA-dependent RNA
8	27	79.4	201	2 F69988	hypothetical prote
9	27	79.4	598	2 B71095	hypothetical prote
10	27	79.4	601	2 D96001	probable 1,4-alpha
11	27	79.4	802	2 T05596	probable potassium
12	27	79.4	1560	2 T30282	calcium-binding pr
13	26	76.5	46	2 F95318	conserved hypothet
14	26	76.5	150	2 B55209	H transfer determi
15	26	76.5	223	2 E84205	deoxyribose-phosph
16	26	76.5	224	2 S02216	deoxyribose-phosph
17	26	76.5	252	2 E72060	conserved hypothet
18	26	76.5	252	2 B86564	CT470 hypothetical
19	26	76.5	269	2 A56487	signal recognition
20	26	76.5	278	2 C86421	Flu18.17 protein -
21	26	76.5	294	2 T45662	hypothetical prote
22	26	76.5	324	2 B84452	probable steroid s
23	26	76.5	490	2 E96010	probable two-compo
24	26	76.5	498	2 H82494	probable NADH dehy
25	26	76.5	533	2 AE1616	ABC transporter (A
26	26	76.5	533	2 AG1253	ABC transporter (A
27	26	76.5	543	2 G83825	hypothetical prote
28	26	76.5	558	2 JC5135	alpha-amylase (EC
29	26	76.5	561	2 S73087	alpha-amylase (EC

30	26	76.5	562	2 G75044	acetolactate synth
31	26	76.5	587	2 D81881	probable succinate
32	26	76.5	587	2 F81138	succinate dehydrog
33	26	76.5	603	2 D70445	aspartate-tRNA lig
34	26	76.5	623	2 B82536	ABC transporter AT
35	26	76.5	1014	2 C83990	beta-galactosidase
36	26	76.5	1085	2 G89056	protein K09H11.3 [
37	25	73.5	136	2 I70179	dynein-like protei
38	25	73.5	152	2 S14388	yabB protein - Esc
39	25	73.5	152	2 E90639	hypothetical prote
40	25	73.5	152	2 E85490	hypothetical prote
41	25	73.5	152	2 AD0067	conserved hypothet
42	25	73.5	152	2 AD0517	acyltransferase li
43	25	73.5	187	2 A86847	dynein-like protei
44	25	73.5	188	2 I70180	dynein-like protei
45	25	73.5	200	2 I70169	dynein-like protei

#### ALIGNMENTS

##### RESULT 1

C97255

thymidine kinase (EC 2.7.1.21) [similarity] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 31-Dec-2001

C:Accession: C97255

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.

; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: C97255

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-195 <KUR>

A:Cross-references: GB:AB001437; PIDN:AAK80830.1; PID:g15025935; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2887

C:Superfamily: thymidine kinase

C:Keywords: phosphotransferase

Query Match 85.3% Score 29; DB 2; Length 195;

Best Local Similarity 62.5%; Pred No. 9;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8

I: I : I I I

DB 130 ELMAIAEF 137

##### RESULT 2

B70728

hypothetical protein RV2558 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: B70728

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70728

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-236 <COL>

A:Cross-references: GB:Z77250; GB:AL123456; NID:g3261617; PIDN:CAB01046.1; PID:e25533

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV2558

Query Match 85.3%; Score 29; DB 2; Length 236;  
 Best Local Similarity 75.0%; Pred. No. 11;  
 Matches 6; Conservative 1; Mismatches 0; Gaps 0;

Oy 1 EVMXVAEF 8  
 ||| ||||  
 Db 216 EVLDVAEF 223

RESULT 3  
 A:Title: hydrolase STY1503 [imported] - Salmonella enterica subsp. enterica serovar Typh  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AJ0673  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 , S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AI0673  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-594 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD01762.1; PID:g16502610; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY1503  
 C:Superfamily: trehalose trehalohydrolase

Query Match 82.4%; Score 28; DB 2; Length 594;  
 Best Local Similarity 75.0%; Pred. No. 52;  
 Matches 6; Conservative 1; Mismatches 0; Gaps 0;

Oy 1 EVMXVAEF 8  
 ||| ||||  
 Db 148 EYMPVAQF 155

RESULT 4  
 AB2042  
 Hypothetical protein all1888 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AB2042  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AB2042  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1256 <KUR>  
 A:Cross-references: GB:BA00000119; PIDN:BA073587.1; PID:g17130978; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: all1888

Query Match 82.4%; Score 28; DB 2; Length 1256;  
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EVMXVAEF 8  
 ||| ||||  
 Db 74 EYMAFAEF 81

RESULT 5

A42545  
 genome polyprotein - Langkat virus (strain TP21) (fragment)  
 N:Contains: nonstructural protein NS1; nonstructural protein NS2a; nonstructural  
 NS5  
 C:Species: Langkat virus  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 19-Jan-2001  
 C:Accession: A42545; A61409; C61409  
 R:Iacono-Connors, L.C.; Schmaljohn, C.S.  
 Virology 188, 875-880, 1992  
 A:Title: Cloning and sequence analysis of the genes encoding the nonstructural p  
 A:Reference number: A42545; MUID:92263794; PMID:1316684  
 A:Accession: A42545  
 A:Molecule type: genomic RNA  
 A:Residues: 1-2638 <IAC>  
 A:Cross-references: GB:S35365; NID:g249315; PIDN:AAB22165.1; PID:g249316  
 R:Guirakhoo, F.; Heinz, F.X.; Mandl, C.W.; Holzmann, H.; Kunz, C.; Gresikova, M.  
 J. Gen. Virol. 72, 333-336, 1991  
 A:Title: The relationship between the flaviviruses Skalica and Langat as reveale  
 A:Reference number: A61409; MUID:91132129; PMID:1847173  
 A:Accession: A61409  
 A:Status: not compared with conceptual translation  
 A:Molecule type: genomic RNA  
 A:Residues: 319-337 <GUI>  
 A:Accession: C61409  
 A:Status: not compared with conceptual translation  
 A:Molecule type: genomic RNA  
 A:Residues: 877-994 <GU2>  
 C:Superfamily: yellow fever virus genome polyprotein  
 C:Keywords: glycoprotein; nonstructural protein; nucleotide binding; P-loop; pol  
 F:1-352/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:353-582/Product: nonstructural protein NS2a #status predicted <N2A>  
 F:583-713/Product: nonstructural protein NS2b #status predicted <N2B>  
 F:714-1334/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:912-919/Region: nucleotide-binding motif A (P-loop)  
 F:1335-1483/Product: nonstructural protein NS4a #status predicted <NAA>  
 F:1484-1735/Product: nonstructural protein NS4b #status predicted <NAB>  
 F:1736-2638/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:85,207,223,873,1212,1671,1950/Binding site: carbohydrate (Aen) (covalent) #sta

Query Match 82.4%; Score 28; DB 1; Length 2638;  
 Best Local Similarity 85.7%; Pred. No. 2.5e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 EVMXVAEF 8  
 ||| ||||  
 Db 153 EMTVAEF 159

RESULT 6  
 T11902  
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain N002 - Thermus aquaticus the  
 C:Species: Thermus aquaticus thermophilus  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 03-Jun-2002  
 C:Accession: T11902  
 R:Iano, T.; Chu, S.S.; Sled, V.S.; Ohnishi, T.; Yagi, T.  
 submitted to the EMBL Data Library, March 1996  
 A:Description: Cloning, sequencing, and expression studies of the proton-transloc  
 A:Reference number: Z17372  
 A:Accession: T11902  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-181 <YAN>  
 A:Cross-references: EMBL:U52917; NID:g1279860; PID:g1279865; PIDN:AAA97942.1  
 C:Genetics:  
 A:Gene: N002  
 C:Keywords: electron transfer; membrane-associated complex; NAD; oxidoreductase

Query Match 79.4%; Score 27; DB 2; Length 181;  
 Best Local Similarity 75.0%; Pred. No. 27;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EVMXVAEF 8  
 ||| ||||

Db 59 EVMGVASE 66

RESULT 7

C69133  
DNA-dependent RNA polymerase, subunit E' - Methanobacterium thermoautotrophicum (strain C:Species: Methanobacterium thermoautotrophicum  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999  
C:Accession: C69133  
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional reference number: A69000; MUID:98037514; PMID:9371463  
A:Accession: C69133  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-184 <MTH>  
A:Cross-references: GB:AF000812; GB:AF000666; NID:g2621298; PIDN:AAB84770.1; PID:g262131  
A:Experimental source: strain Delta H  
C:Genetics:  
A:Gene: MTH264  
A:Start codon: TTG  
C:Superfamily: DNA-directed RNA polymerase subunit E

Query Match 79.4%; Score 27; DB 2; Length 184;  
Best Local Similarity 62.5%; Pred. No. 27;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVMXVAEF 8  
||: ||||  
Db 88 EVIEIAEF 95

RESULT 8

F69988  
hypothetical protein ytbQ - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
C:Accession: F69988  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertoni, A.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chao, A.; Ehrlich, S.D.; Emmons, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Patro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yanamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: F69988  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-201 <KUN>  
A:Cross-references: GB:Z99119; GB:AL009126; NID:g2635411; PIDN:CAB14996.1; PID:eil185891;  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ytbQ

Query Match 79.4%; Score 27; DB 2; Length 201;  
Best Local Similarity 50.0%; Pred. No. 30;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVMXVAEF 8  
||: ||||  
Db 22 DIMDIAEF 29

RESULT 9

B71095  
hypothetical protein PH1023 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000  
C:Accession: B71095  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A:Reference number: A71000; MUID:98344137; PMID:9679194  
A:Accession: B71095  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-598 <KAW>  
A:Cross-references: GB:AF000004; NID:g3236131; PIDN:BAA30120.1; PID:g3257437  
A:Experimental source: strain OT3  
A:Note: This accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH1023

Query Match 79.4%; Score 27; DB 2; Length 598;  
Best Local Similarity 62.5%; Pred. No. 93;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVMXVAEF 8  
||: ||||  
Db 280 EIVNVAEF 287

RESULT 10

D96001  
probable 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) [imported] - Sinorhizobium m  
C:Species: Sinorhizobium meliloti  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 14-Sep-2001.  
C:Accession: D96001  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Her  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A:Title: The complete sequence of the 1,683-Kb pSymB megaplasmid from the N2-fixing e  
A:Reference number: A95842; MUID:21396508; PMID:11481431  
A:Accession: D96001  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-601 <KUR>  
A:Cross-references: GB:AL591985; PIDN:CAC49676.1; PID:g15141163; GSPDB:GN00167  
A:Experimental source: strain 1021, megaplasmid pSymB  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl  
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau  
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: glgB2; SMB21447  
A:Genome: plasmid  
C:Superfamily: trehalose trehalohydrolase  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 79.4%; Score 27; DB 2; Length 601;  
Best Local Similarity 62.5%; Pred. No. 94;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVMXVAEF 8  
||: ||||  
Db 151 EIMPVAQF 158

RESULT 11

## T05596

probable potassium transport protein F9D16.110 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 18-Aug-2000  
 C:Accession: T05596  
 R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.  
 submitted to the Protein Sequence Database, February 1999  
 A:Reference number: Z15419  
 A:Accession: T05596  
 A:Molecule type: DNA  
 A:Residues: 1-802 <BEV>  
 A:Cross-references: EMBL:AL035394  
 A:Experimental source: cultivar Columbia; BAC clone F9D16  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 1/3; 16/3; 102/1; 185/1; 209/1; 296/1; 313/3; 352/1; 437/1  
 A:Note: F9D16.110  
 C:Superfamily: barley probable potassium transport protein HAK1  
 C:Keywords: ion transport

Query Match 79.4%; Score 27; DB 2; Length 802;  
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VMXVAEF 8  
 || :|||  
 Db 629 VMSIAEF 635

## RESULT 12

T30282  
 calcium-binding protein - sea urchin (Strongylocentrotus purpuratus)  
 C:Species: Strongylocentrotus purpuratus (purple urchin)  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
 C:Accession: T30282  
 R:Soltyk-Espanola, M.; Klinzing, D.C.; Pfarr, K.; Burke, R.D.; Ernst, S.G.  
 Dev. Biol. 165, 73-85, 1994  
 A:Title: Endo16, a large multidomain protein found on the surface and ECM of endodermal  
 A:Reference number: Z20805; MUID:94374583; PMID:8088452  
 A:Accession: T30282  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1560 <SOL>  
 A:Cross-references: EMBL:L34680; NID:G511893; PID:G511894; PIDN:AAA30047.1  
 C:Genetics:  
 A:Note: endo16

Query Match 79.4%; Score 27; DB 2; Length 1560;  
 Best Local Similarity 62.5%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVMXVAEF 8  
 || :|||  
 Db 705 ELLVAEF 712

## RESULT 13

F95318  
 conserved hypothetical protein Sma0833 [imported] - Sinorhizobium meliloti (strain 1021)  
 C:Species: Sinorhizobium meliloti  
 C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C:Accession: F95318  
 R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows  
 ; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti  
 A:Reference number: A95262; MUID:21396509; PMID:11481432  
 A:Accession: F95318  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-46 <KUR>  
 A:Cross-references: GB:AE006469; PIDN:AAK65112.1; PID:G14523550; GSPDB:GN00165  
 A:Experimental source: strain 1021, megaplasmid pSymA

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-  
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher,  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.;  
 hehault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.  
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: Sma0833  
 A:Genome: plasmid

Query Match 76.5%; Score 26; DB 2; Length 46;  
 Best Local Similarity 71.4%; Pred. No. 11;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVMXVAE 7  
 || :|||  
 Db 17 EIMVAE 23

## RESULT 14

B55209  
 H transfer determinant A - plasmid R27  
 C:Species: plasmid R27  
 C:Date: 05-May-1995 #sequence\_revision 05-May-1995 #text\_change 05-Nov-1999  
 C:Accession: B55209  
 R:Whelan, K.F.; Maher, D.; Colleran, E.; Taylor, D.E.  
 J. Bacteriol. 176, 2242-2251, 1994  
 A:Title: Genetic and nucleotide sequence analysis of the gene htdA, which regula  
 A:Reference number: A55209; MUID:94209223; PMID:7908903  
 A:Accession: B55209  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-150 <WHE>  
 A:Cross-references: GB:L20342; NID:G410304; PIDN:AA00506.1; PID:G1326032  
 C:Genetics:  
 A:Gene: htdA  
 A:Genome: plasmid

Query Match 76.5%; Score 26; DB 2; Length 150;  
 Best Local Similarity 50.0%; Pred. No. 39;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVMXVAEF 8  
 || :|||  
 Db 18 EVLSISEF 25

## RESULT 15

E64205  
 deoxyribose-phosphate aldolase (EC 4.1.2.4) - Mycoplasma genitalium  
 C:Species: Mycoplasma genitalium  
 C:Date: 17-Nov-1995 #sequence\_revision 17-Nov-1995 #text\_change 21-Jul-2000  
 C:Accession: E64205  
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischman  
 ; C.A.; Venter, J.C.  
 Science 270, 397-403, 1995  
 A:Title: The minimal gene complement of Mycoplasma genitalium.  
 A:Reference number: A64200; MUID:96026346; PMID:7569993  
 A:Accession: E64205  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-223 <TIGR>  
 A:Cross-references: GB:U039684; GB:L43967; NID:G3844650; PIDN:AA071266.1; PID:G10

A:Experimental source: strain G-37  
 C:Genetics:  
 A:Genetic code: SGC3  
 C:Superfamily: deoxyribose-phosphate aldolase  
 C:Keywords: aldehyde-lyase; carbon-carbon lyase



Query Match 76.5%; Score 26; DB 2; Length 223;  
Best Local Similarity 71.4%; Pred. No. 59;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VMXVAEF 8  
||:||||  
Db 93 VMNTAEF 99

Search completed: April 23, 2003, 12:36:28  
Job time : 11.6552 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:33:40 ; Search time 5.24138 Seconds  
(without alignments)  
63.306 Million cell updates/sec

Title: US-09-723-722A-81  
Perfect score: 34  
Sequence: 1 EVMXVAEF 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match %	Length	ID	Description
1	29	85.3	236	YP58_MYCTU	Q50740 mycobacteri
2	28	82.4	3414	POUG_LANVT	P29837 l genome po
3	27	79.4	181	YQ02_THETH	Q56221 thermus the
4	27	79.4	253	NTBQ_BACSU	P53560 bacillus su
5	26	76.5	223	DEOC_MYCGE	P47296 mycoplasma
6	26	76.5	224	DEOC_MYCPN	P09924 mycoplasma
7	26	76.5	252	RECO_CHLPN	Q92745 chlamydia p
8	26	76.5	269	SRPB_MOUSE	P47758 mus musculu
9	26	76.5	271	SRPB_HUMAN	Q95588 homo sapien
10	26	76.5	457	ARLY_PASMU	P57909 pasteurella
11	26	76.5	603	SYD_AQUAE	Q67589 aquifex aeo
12	26	76.5	1014	BGAL_BACHD	Q9K9C6 bacillus ha
13	25	73.5	152	MRAZ_ECO57	Q8X923 escherichia
14	25	73.5	152	MRAZ_ECOLI	P22186 escherichia
15	25	73.5	152	MRAZ_SALTI	Q8Z9H5 salmonella
16	25	73.5	152	MRAZ_SALTY	Q8Z9U9 salmonella
17	25	73.5	152	MRAZ_VERPE	Q8ZIF8 versinia pe
18	25	73.5	234	RECO_CHLMU	Q9PIS3 chlamydia m
19	25	73.5	413	2ABB_RABIT	Q00006 o serine/th
20	25	73.5	426	2ABA_PIG	Q29090 s serine/th
21	25	73.5	443	2ABB_HUMAN	Q00005 homo sapien
22	25	73.5	443	2ABB_PIG	P54614 sus scrofa
23	25	73.5	443	2ABB_RAT	P36877 r serine/th
24	25	73.5	447	2ABA_HUMAN	Q00007 h serine/th
25	25	73.5	447	2ABA_RAT	P36876 r serine/th
26	25	73.5	453	2ABD_RAT	P56932 r serine/th
27	25	73.5	499	2ABA_DROME	P36872 drosophila
28	25	73.5	513	BMF6_HUMAN	P22004 homo sapien
29	25	73.5	632	ETFD_SCHPO	P87111 s probable
30	25	73.5	808	FGR4_MOUSE	Q03142 mus musculu
31	25	73.5	4466	DYHC_ANTCR	P39057 anthocidari
32	25	73.5	4466	DYHC_TRIGR	P23098 tripneustes
33	25	73.5	4486	DYH9_HUMAN	Q9nyc9 homo sapien

34	24	70.6	91	1	YAH0_ECOLI	P75694 escherichia
35	24	70.6	145	1	RL13_HALMA	P29198 haloarcula
36	24	70.6	148	1	CYC6_CHLRE	P08197 chlamydomon
37	24	70.6	157	1	RAP_TAROF	O49065 taraxacum o
38	24	70.6	186	1	RS7_METTH	O27130 methanobact
39	24	70.6	186	1	YA08_PYRHO	O58736 pyrococcus
40	24	70.6	187	1	RPE1_METJA	Q57840 methanococ
41	24	70.6	263	1	LPXA_CAUCR	Q9A715 caulobacter
42	24	70.6	263	1	NIH2_METIV	P08624 methanobact
43	24	70.6	291	1	AMP4_ARCFU	O28438 archaeoglob
44	24	70.6	295	1	SUOE_BOVIN	P19217 bos taurus
45	24	70.6	298	1	YSMK_CAEEL	Q19408 caenorhabdi

ALIGNMENTS

RESULT 1  
YP58\_MYCTU STANDARD; PRT; 236 AA.  
AC Q50740:  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein RV2558.  
GN RV2558 OR MT2635 OR MTCY9C4.10C.  
OS Mycobacterium tuberculosis  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37Rv;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.F., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
Rutler S., Seeger K., Skelton S., Squares S., Squares R.,  
Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: SPONG, TO M.TUBERCULOSIS RV2557.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC EMBL; 277250; CAB01046.1;  
DR EMBL; AE007098; AAK46947.1;  
DR TIGR; MT2635;  
DR Tuberculist; RV2558;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 236 AA; 25718 MW; 13E3B049D8F79C6B\_CRC64;

Query Match 85.3%; Score 29; DB 1; Length 236;  
 Best Local Similarity 75.0%; Pred. No. 4.9;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVMXVAEF 8  
 ||: ||||  
 Db 216 EVLDVAEF 223

## RESULT 2

POLG\_LANVT STANDARD; PRT; 3414 AA.  
 AC P29837;  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein);  
 DE Envelope protein E; Matrix protein (Envelope protein M); Major  
 DE NS2A; Nonstructural protein NS2B; Helicase/protease (EC 3.4.21.98)  
 DE (NS3); Nonstructural protein NS4A; Nonstructural protein NS4B; RNA-  
 DE directed RNA polymerase (EC 2.7.7.48) (NS5)].  
 OS Langkat virus (strain TP21).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Flavivirus.  
 OX NCBI\_TaxID=31638;  
 RN [1]  
 RP SEQUENCE OF 1-776 FROM N.A.  
 RX MEDLINE=92074260; PubMed=1720591;  
 RA Mandl C.W., Iacono-Connors L., Wallner G., Holzmann H., Kunz C.,  
 RA Heinz F.X.;  
 RT "Sequence of the genes encoding the structural proteins of the low-  
 RT virulence tick-borne flaviviruses Langkat TP21 and Yellantsev.";  
 RL Virology 185:891-895(1991).  
 [2]  
 RP SEQUENCE OF 777-3414 FROM N.A.  
 RX MEDLINE=92263794; PubMed=1316684;  
 RA Iacono-Connors L.C., Schmaljohn C.S.;  
 RT "Cloning and sequence analysis of the genes encoding the  
 RT nonstructural proteins of Langkat virus and comparative analysis with  
 RT other flaviviruses.";  
 RL Virology 188:875-880(1992).  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the p6  
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC [RNA](N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M73835; AAA02740.1; ALT\_TERM.  
 CC EMBL; S35365; AAB22165.1; -.  
 CC PIR; A41704; A41704.  
 CC PIR; A42545; A42545.  
 CC HSP; P14336; 1SVB.  
 CC MEROPS; S07.001; -.  
 CC InterPro; IPR001410; DEAD.  
 CC InterPro; IPR000069; Flavi\_M.  
 CC InterPro; IPR001157; Flavi\_NS1.  
 CC InterPro; IPR000752; Flavi\_NS2A.  
 CC InterPro; IPR000487; Flavi\_NS2B.  
 CC InterPro; IPR000404; Flavi\_NS4A.  
 CC InterPro; IPR001528; Flavi\_NS4B.

DR InterPro; IPR000208; Flavi\_NS5.  
 DR InterPro; IPR001122; Flavi\_capsidC.  
 DR InterPro; IPR000336; Flavi\_glycoprote.  
 DR InterPro; IPR001850; Flavi\_helicase.  
 DR InterPro; IPR002535; Flavi\_propep.  
 DR InterPro; IPR002877; FtsJ.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00869; Flavi\_glycoprot; 1.  
 DR Pfam; PF00948; Flavi\_helicase; 1.  
 DR Pfam; PF00949; Flavi\_helicase; 1.  
 DR Pfam; PF00972; Flavi\_NS5; 1.  
 DR Pfam; PF01002; Flavi\_NS2B; 1.  
 DR Pfam; PF01003; Flavi\_capsid; 1.  
 DR Pfam; PF01004; Flavi\_M; 1.  
 DR Pfam; PF01005; Flavi\_NS2A; 1.  
 DR Pfam; PF01349; Flavi\_NS4B; 1.  
 DR Pfam; PF01350; Flavi\_NS4B; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 DR Pfam; PF01728; FtsJ; 1.  
 DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
 DR ProDom; PD001496; Flavi\_NS1; 1.  
 DR ProDom; PD001556; Flavi\_glycoprote; 1.  
 DR SMART; SM00490; HELIC\_C; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;  
 KW ATP-binding; Transmembrane; Nonstructural protein.  
 FT INIT\_MET 1 1  
 FT CHAIN 1 112  
 FT CHAIN 113 205  
 FT CHAIN 206 280  
 FT CHAIN 281 776  
 FT CHAIN 777 1128  
 FT CHAIN 1129 1358  
 FT CHAIN 1359 1489  
 FT CHAIN 1490 2110  
 FT CHAIN 2111 2259  
 FT CHAIN 2260 2511  
 FT CHAIN 2512 3414  
 FT NP\_BIND 1688 1695  
 FT SITE 1779 1782  
 FT TRANSMEM 103 119  
 FT TRANSMEM 262 278  
 FT TRANSMEM 728 744  
 FT TRANSMEM 758 774  
 FT DISULFID 283 310  
 FT DISULFID 340 396  
 FT DISULFID 354 385  
 FT DISULFID 372 401  
 FT DISULFID 466 570  
 FT DISULFID 587 618  
 FT CARBOHYD 144 144  
 FT CARBOHYD 434 434  
 SQ SEQUENCE 3414 AA; 378017 MW; 59CB7E95DD70D82E CRC64;  
 Query Match 82.4%; Score 28; DB 1; Length 3414;  
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Caps 0;  
 Qy 2 VMXVAEF 8  
 ||: ||||  
 Db 929 VMTVAEF 935  
 RESULT 3  
 ID NOO2\_THETH STANDARD; PRT; 181 AA.  
 AC O56221;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3) (NADH  
 DN dehydrogenase 1, chain 2) (NDH-1, chain 2).  
 GN NQ02.  
 OS Thermus thermophilus.  
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Thermales;  
 CC Thermaceae; Thermus.  
 OX NCBI\_TaxID=274;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN-HB8 / ATCC 27634;  
 RX MEDLINE=97172490; PubMed=9020134;  
 RA Yano T., Chu S.S., Sled V.D., Ohnishi T., Yagi T.;  
 RT "The proton-translocating NADH-quinone oxidoreductase (NDH-1) of  
 thermophilic bacterium Thermus thermophilus HB-8. Complete DNA  
 RT sequence of the gene cluster and thermostable properties of the  
 RT expressed NQ02 subunit".  
 RL J. Biol. Chem. 272:4201-4211(1997).  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -1- COFACTOR: BINDS A 2FE-2S CLUSTER.  
 CC -1- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS.  
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 24 kDa SUBUNIT FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U52917; AAA97942.1; -  
 DR InterPro; IPR002023; Cmplx1\_24kDa.  
 DR Pfam; PF01257; complex1\_24kd; 1.  
 DR ProDom; PD003859; Cmplx1\_24kDa; 1.  
 DR PROSITE; PS01099; COMPLEX1\_24K; 1.  
 KW Oxidoreductase; NAD; Ubiquinone; Iron-sulfur.  
 FT METAL 83 83 IRON-SULFUR (2FE-2S) (POTENTIAL).  
 FT METAL 88 88 IRON-SULFUR (2FE-2S) (POTENTIAL).  
 FT METAL 124 124 IRON-SULFUR (2FE-2S) (POTENTIAL).  
 FT METAL 128 128 IRON-SULFUR (2FE-2S) (POTENTIAL).  
 SQ SEQUENCE 181 AA; 20286 MW; 484FE09245C613BE CRC64;  
 Query Match 79.4%; Score 27; DB 1; Length 181;  
 Best Local Similarity 75.0%; Pred. No. 12;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 EVMXVAEF 8  
 ||| |||  
 Db 59 EVMGVASF 66  
 RESULT 4  
 ID YTBQ\_BACSU STANDARD; PRT; 253 AA.  
 AC PS3560; O34914;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein ytbQ.  
 GN YTBQ.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96312354; PubMed=8763940;  
 RA Bower S., Perkins J.B., Yocum R.R., Howitt C.L., Rahaim P.,  
 RA Pero J.;  
 RT "Cloning, sequencing, and characterization of the Bacillus subtilis  
 RT biotin biosynthetic operon".  
 RL J. Bacteriol. 178:4122-4130(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Chouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Ghim S.Y., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Guiseppi G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauei C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis".  
 RL Nature 390:249-256(1997).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U51868; AAB17463.1; -  
 DR EMBL; AF008220; AAC00267.1; ALT\_INIT.  
 DR EMBL; Z99119; CAB14996.1; ALT\_INIT.  
 DR Subtilist; BG11787; ytbQ.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 253 AA; 28233 MW; 572CC4681F04C650 CRC64;  
 Query Match 79.4%; Score 27; DB 1; Length 253;  
 Best Local Similarity 50.0%; Pred. No. 17;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 EVMXVAEF 8  
 :|| :|||  
 Db 74 DIMDIAEF 81  
 RESULT 5  
 ID DEOC\_MYCGE STANDARD; PRT; 223 AA.  
 AC P47296;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyriboaldolase)  
 DE (Deoxyriboaldolase).  
 GN DEOC OR MG050.  
 OS Mycoplasma genitalium.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=2097;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 33530 / G-37;  
 RX MEDLINE=96026346; PubMed=7569993;  
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
 RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
 RA Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,  
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,  
 RT "The minimal gene complement of Mycoplasma genitalium";  
 RL Science 270:397-403(1995).  
 CC -1- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate - D-  
 CC glyceraldehyde 3-phosphate + acetaldehyde.  
 CC -1- PATHWAY: Nucleotide and deoxyribonucleotide catabolism.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE DEOC/FBAB FAMILY OF ALDOLASES.  
 CC DEOC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U39684; AAC71266.1; -;  
 DR TIGR: MG050;  
 DR InterPro: IPR002915; Deoc.  
 DR InterPro: IPR003009; FNN\_enzyme.  
 DR Pfam: PF01791; Deoc; 1.  
 KW Lyase; Schiff base; Complete proteome.  
 FT BINDING 152 152 SCHIFF-BASE (BY SIMILARITY).  
 SQ SEQUENCE 223 AA; 24675 MW; 332430231CE99DB0 CRC64;  
 -----  
 Query Match 76.5%; Score 26; DB 1; Length 223;  
 Best Local Similarity 71.4%; Pred. No. 26;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 VMXVAEF 8  
 II :III  
 DB 93 VMNIAEF 99  
 -----  
 RESULT 6  
 DEOC\_MYCPN STANDARD; PRT; 224 AA.  
 AC P09924;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyriboaldolase)  
 DE (deoxyriboaldolase).  
 GN DEOC OR MPN063 OR MP091.  
 OS Mycoplasma pneumoniae.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=2104;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 29342 / M129;  
 RX MEDLINE=89128453; PubMed=2492658;  
 RA Loechel S., Inamine J.M., Hu P.-C.;  
 RT "Nucleotide sequence of the deoc gene of Mycoplasma pneumoniae";  
 RL Nucleic Acids Res. 17:801-801(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 29342 / M129;  
 RX MEDLINE=97105885; PubMed=8948633;  
 RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,  
 RA Herrmann R.;

RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
 pneumoniae";  
 RL Nucleic Acids Res. 24:4420-4449(1996).  
 CC -1- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate - D-  
 CC glyceraldehyde 3-phosphate + acetaldehyde.  
 CC -1- PATHWAY: Nucleotide and deoxyribonucleotide catabolism.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE DEOC/FBAB FAMILY OF ALDOLASES.  
 CC DEOC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X13544; CAA31897.1; -;  
 DR EMBL: AE000011; AAB95739.1; -;  
 DR PIR: S02216; S02216.  
 DR InterPro: IPR002915; Deoc.  
 DR Pfam: PF01791; Deoc; 1.  
 KW Lyase; Schiff base; Complete proteome.  
 FT BINDING 152 152 SCHIFF-BASE (BY SIMILARITY).  
 SQ SEQUENCE 224 AA; 24878 MW; 73C3E4932E7881F7 CRC64;  
 -----  
 Query Match 76.5%; Score 26; DB 1; Length 224;  
 Best Local Similarity 71.4%; Pred. No. 26;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 VMXVAEF 8  
 II :III  
 DB 93 VMNIAEF 99  
 -----  
 RESULT 7  
 RECO\_CHLPN STANDARD; PRT; 252 AA.  
 AC Q927W5;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE DNA repair protein reco (Recombination protein O).  
 GN RECO OR CPN0589 OR CP0159 OR CPJ0589.  
 OS Chlamydia pneumoniae (Chlamydophila pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 pneumoniae AR39";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,

RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CWL029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -!- FUNCTION: Involved in DNA repair and recF pathway recombination  
 CC (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE RECO FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE001643; AAD18728.1; -  
 DR EMBL: AE002177; AAF38039.1; -  
 DR EMBL: AP002547; BAA98796.1; -  
 DR TIGR: CP0159; -  
 DR InterPro: IPR003717; RecO.  
 DR Pfam: PF02565; RecO; 1.  
 KW DNA repair; DNA recombination; Complete proteome.  
 SQ SEQUENCE 252 AA; 28953 MW; 7207B63D9D352460 CRC64;  
 -----  
 Query Match 76.5%; Score 26; DB 1; Length 252;  
 Best Local Similarity 50.0%; Pred. No. 30;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 EVNXYAEF 8  
 I:::III  
 DB 202 ELLAIAEF 209  
 -----  
 RESULT 8  
 SRPB\_MOUSE STANDARD: PRT; 269 AA.  
 ID SRPB\_MOUSE  
 AC P47758; Q9D872;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Signal recognition particle receptor beta subunit (SR-beta).  
 GN SRPB.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95146535; PubMed=7844142;  
 RA Miller J.D., Tajima S., Lauffer L., Walter P.;  
 RT "The beta subunit of the signal recognition particle receptor is a  
 RT transmembrane GTPase that anchors the alpha subunit, a peripheral  
 RT membrane GTPase, to the endoplasmic reticulum membrane."  
 RL J. Cell Biol. 128:273-282(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Small intestine;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,  
 RA Sakimi L.M., Staubli R., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Schirali J., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hall D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 CC -!- FUNCTION: HAS GTPASE ACTIVITY. MAY MEDIATE THE MEMBRANE  
 CC ASSOCIATION OF SR ALPHA.  
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.  
 CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum membrane.  
 CC -!- SIMILARITY: TO OTHER SRP BETA SUBUNITS; DISTANTLY RELATED TO RAS  
 CC SUPERFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U17343; AAA69976.1; -  
 DR EMBL: AK008383; BAB25638.1; -  
 DR MGI: 102964; Srprb.  
 DR InterPro: IPR003575; Small\_GTPase.  
 DR SMART: SM00010; small\_GTPase; 1.  
 KW Signal recognition particle; Transmembrane; Receptor;  
 KW Endoplasmic reticulum; GTP-binding.  
 FT TRANSMEM 35 55 POTENTIAL.  
 FT NP\_BIND 69 76 GTP (POTENTIAL).  
 FT NP\_BIND 115 119 GTP (POTENTIAL).  
 FT NP\_BIND 178 181 GTP (POTENTIAL).  
 FT CONFLICT 16 16 A -> P (IN REF. 2).  
 FT CONFLICT 173 173 L -> P (IN REF. 1; AA SEQUENCE).  
 SQ SEQUENCE 269 AA; 29579 MW; 041175FA6891DA37 CRC64;  
 -----  
 Query Match 76.5%; Score 26; DB 1; Length 269;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 EVNXYAEF 8  
 I:::III  
 DB 149 EVKDVAEF 156  
 -----  
 RESULT 9  
 SRPB\_HUMAN STANDARD: PRT; 271 AA.  
 ID SRPB\_HUMAN  
 AC Q9Y5M8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Signal recognition particle receptor beta subunit (SR-beta) (Protein  
 DE APMCF1).  
 GN SRPB.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast carcinoma;  
 RA Yan W., Zhu F., Chai Y., Zhao Z., Li Q., Wang C.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
 RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
 RA Ninomiya K., Iwayanagi T.;

RT "NEDO human cDNA sequencing project.";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: HAS GTPASE ACTIVITY. MAY MEDIATE THE MEMBRANE  
 CC ASSOCIATION OF SR ALPHA (By similarity).  
 CC -1- SUBUNIT: HETERO-DIMER OF AN ALPHA AND A BETA CHAIN (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum membrane (By  
 CC similarity).  
 CC -1- SIMILARITY: TO OTHER SRP BETA SUBUNITS; DISTANTLY RELATED TO RAS  
 CC SUPERFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF141882; AAD34888.2; -;  
 DR EMBL; AK027525; BAB55176.1; -;  
 KW Signal recognition particle; Transmembrane; Receptor;  
 KW Endoplasmic reticulum; GTP-binding.  
 FT TRANSMEM 37 57 POTENTIAL.  
 FT NP\_BIND 71 78 GTP (POTENTIAL).  
 FT NP\_BIND 117 121 GTP (POTENTIAL).  
 FT NP\_BIND 180 183 GTP (POTENTIAL).  
 FT CONFLICT 9 9 V -> L (IN REF. 2).  
 SQ SEQUENCE 271 AA; 29651 MW; 218CALFAFE00BED3 CRC64;  
 Query Match 76.5%; Score 26; DB 1; Length 271;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 EVMXVAEF 8  
 II IIII  
 Db 151 EVKDVAF 158  
 RESULT 10  
 ARLY\_PASMU STANDARD; PRT; 457 AA.  
 AC P57909;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Argininosuccinate lyase (EC 4.3.2.1) (Argininosuccinase) (ASAL).  
 GN ARGH OR Pm120.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Pasteurella.  
 OX NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Pm70;  
 RX MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 CC -1- CATALYTIC ACTIVITY: N-(L-arginino)succinate -> fumarate + L-  
 CC arginine.  
 CC -1- PATHWAY: Arginine biosynthesis; last step.  
 CC -1- SIMILARITY: BELONGS TO THE LYASE 1 FAMILY. ARGININOSUCCINATE LYASE  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE006153; AAK03204.1; -;

DR HSSP; P04424; IAOS.  
 DR InterPro: IPR003031; D\_crystallin.  
 DR InterPro: IPR000362; Fumarate\_lyase.  
 DR Pfam: PF00206; lyase\_1; 1.  
 DR PRINTS: PR00149; FUMRATEDIASE.  
 DR TIGRFAMS: TIGR00838; argH; 1.  
 DR PROSITE: PS00163; FUMARATE\_LYASES; 1.  
 KW Arginine biosynthesis; Lyase; Complete proteome.  
 SQ SEQUENCE 457 AA; 50709 MW; 936CC039B3BDBF6A CRC64;  
 Query Match 76.5%; Score 26; DB 1; Length 457;  
 Best Local Similarity 62.5%; Pred. No. 54;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 EVMXVAEF 8  
 I : IIII  
 Db 403 EALSVAEF 410  
 RESULT 11  
 SYD\_AQUAE STANDARD; PRT; 603 AA.  
 ID SYD\_AQUAE  
 AC O67589;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Aspartyl-uracil synthetase (EC 6.1.1.12) (Aspartate--trRNA ligase)  
 DE (ASPRS).  
 DE ASPRS OR AQ\_1677.  
 GN ASPRS OR AQ\_1677.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;  
 OC Aquifex.  
 OX NCBI\_TaxID=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RX MEDLINE=98196666; PubMed=9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujaay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RT aeolicus.";  
 RL Nature 392:353-358(1998).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-aspartate + trRNA(Asp) -> AMP +  
 CC diphosphate + L-aspartyl-trRNA(Asp).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000750; AAC07548.1; -;  
 DR HSSP; P36419; LEFW.  
 DR InterPro: IPR002106; AATRNA\_ligaseII.  
 DR InterPro: IPR004524; Asps\_bact.  
 DR InterPro: IPR004115; GAD\_dom.  
 DR InterPro: IPR004364; trRNA-synt\_2.  
 DR InterPro: IPR002312; trRNA-synt\_asp.  
 DR InterPro: IPR004365; trRNA\_anti.  
 DR Pfam: PF00152; trRNA-synt\_2; 2.  
 DR Pfam: PF01336; trRNA\_anti; 1.  
 DR Pfam: PF02938; GAD; 1.  
 DR PRINTS: PR01042; TRNASYNTHASP.  
 DR TIGRFAMS: TIGR00459; asps\_bact; 1.  
 DR PROSITE: PS50862; AA\_TRNA\_LIGASE\_II; 1.  
 KW Aminoacyl-trRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Complete proteome.



SQ SEQUENCE 603 AA; 69729 MW; 9DFFBD840C8DCC1C CRC64;

Query Match 76.5%; Score 26; DB 1; Length 603;  
Best Local Similarity 85.7%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAE 7

Db 254 EVMXVAE 260

RESULT 12

BGAL\_BACHD STANDARD; PRT; 1014 AA.

AC Q9K9C6;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Beta-galactosidase (EC 3.2.1.23) (Lactase).

GN LACZ OR BH2723.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=85665;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RX MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
Horikoshi K.;

RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
halodurans and genomic sequence comparison with Bacillus subtilis.";

RL Nucleic Acids Res. 28:4317-4331(2000).

CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-  
galactose residues in beta-D-galactosides.

CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AP001516; BAB06442.1; -

DR HSSP; P00722; 1BGL.

DR InterPro; IPR004200; Bgal\_small\_C.

DR InterPro; IPR004199; Bgal\_small\_N.

DR InterPro; IPR001649; GH\_2.

DR Pfam; PF02930; Bgal\_small\_C; 1.

DR Pfam; PF02929; Bgal\_small\_N; 1.

DR Pfam; PF00703; Glyco\_hydro\_2; 1.

DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.

DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.

DR PRINTS; PR00132; GLYDRLASE2.

DR PROSITE; PS00719; GLYCOSYL HYDROL\_F2.1; 1.

DR PROSITE; PS00608; GLYCOSYL HYDROL\_F2.2; 1.

KW Hydrolase; Glycosidase; Complete proteome.

FT ACT\_SITE 460 460 PROTON DONOR (BY SIMILARITY).

FT ACT\_SITE 527 527 NUCLEOPHILE (BY SIMILARITY).

SQ SEQUENCE 1014 AA; 116656 MW; 30405EB697C72798 CRC64;

Query Match 76.5%; Score 26; DB 1; Length 1014;

Best Local Similarity 85.7%; Pred. No. 1.2e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAE 7

Db 253 EVMXVAE 259

RESULT 13

MRAZ\_ECO57 STANDARD; PRT; 152 AA.  
ID MRAZ\_ECO57  
AC Q8Y9Z3;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Protein mraz.  
GN MRAZ OR Z0091 OR ECS0085.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";

RL Nature 409:529-533(2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12.";

RL DNA Res. 8:11-22(2001).

CC -1- SIMILARITY: BELONGS TO THE MRAZ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF005185; AAG54385.1; -

DR EMBL; AP002550; BAB33508.1; -

DR InterPro; IPR003444; UPF0040.

DR Pfam; PF02381; UPF0040; 2.

KW Complete proteome.

SQ SEQUENCE 152 AA; 17360 MW; 48277D972BEDC765 CRC64;

Query Match 73.5%; Score 25; DB 1; Length 152;

Best Local Similarity 62.5%; Pred. No. 32;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8

Db 107 EVMXVGF 114

RESULT 14

MRAZ\_ECOLI

ID MRAZ\_ECOLI STANDARD; PRT; 152 AA.

AC P22186;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Protein mraz.

GN MRAZ OR B0081.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

```
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90251464; PubMed=2187182;
RA Gomez M.J., Fluoret B., van Heijenoort J., Ayala J.A.;
RT "Nucleotide sequence of the regulatory region of the gene pbpB of
RN Escherichia coli.";
RL Nucleic Acids Res. 18:2813-2813(1990).
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -!- SIMILARITY: BELONGS TO THE MRZ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52063; CAA36284.1; -
CC EMBL; X55034; CAA38858.1; -
CC EMBL; AE000118; AAC73192.1; -
CC PIR; S14388; S14388.
CC EcoGene; EG11084; mraz.
CC InterPro; IPR003444; UPF0040.
CC Pfam; PF02381; UPF0040; 2.
CC TIGRFAMs; TIGR00242; TIGR00242; 1.
CC Complete proteome.
CC SEQUENCE 152 AA; 17386 MW; 3EE1A6FA9D2B1C01 CRC64;
Query Match 73.5%; Score 25; DB 1; Length 152;
Best Local Similarity 62.5%; Pred. No. 32;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 EVMXVAEF 8
Db 107 EVMLVQGF 114
[1] | | | | |
[1] | | | | |

RESULT 15
MRZ_SALTI STANDARD; PRT; 152 AA.
AC Q829H5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein mraz.
GN MRZ OR STY0139
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
```

```
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
CC -!- SIMILARITY: BELONGS TO THE MRZ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL627265; CAD01276.1; -
CC InterPro; IPR003444; UPF0040.
CC Pfam; PF02381; UPF0040; 2.
CC TIGRFAMs; TIGR00242; TIGR00242; 1.
CC Complete proteome.
CC SEQUENCE 152 AA; 17417 MW; B76922CA24F477C8 CRC64;
Query Match 73.5%; Score 25; DB 1; Length 152;
Best Local Similarity 62.5%; Pred. No. 32;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 EVMXVAEF 8
Db 107 EVMLVQGF 114
[1] | | | | |
[1] | | | | |
```

Search completed: April 23, 2003, 12:35:45  
Job time : 5.24138 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: April 23, 2003, 12:33:40 ; Search time 34.2069 Seconds  
(without alignments)  
48.188 Million cell updates/sec

Title: US-09-723-722a-81

Perfect score: 34

Sequence: 1 EVMXVAEF 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp-organelle.\*

9: sp-phage.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	30	88.2	284	17	Q8u3z7 pyrococcus
2	29	85.3	195	16	Q97f65 clostridium
3	28	82.4	262	17	Q8u0a9 pyrococcus
4	28	82.4	594	16	Q8zpf0 salmonella
5	28	82.4	594	16	Q8z724 salmonella
6	28	82.4	622	16	Q8xt78 ralstonia
7	28	82.4	776	10	Q94il7 populus tre
8	28	82.4	1256	16	Q8yvt6 anabaena sp
9	28	82.4	3414	12	Q9ig40 langat viru
10	28	82.4	3414	12	Q9ig39 langat viru
11	27	79.4	184	17	O26364 methanobact
12	27	79.4	338	17	Q9hl69 thermoplasm
13	27	79.4	598	17	O58774 pyrococcus
14	27	79.4	601	16	Q92u63 rhizobium m
15	27	79.4	775	10	Q9fe38 arabidopsis
16	27	79.4	802	10	Q9sur2 arabidopsis

17	79.4	814	10	Q8VXQ3	Q8vxq3 cymodocea n
18	79.4	1158	5	Q96594	Q965s4 caenorhabdi
19	79.4	1560	5	Q26644	Q26644 strongyloce
20	76.5	46	16	Q922L0	Q922l0 rhizobium m
21	76.5	92	9	Q9FZT5	Q9fzt5 pseudomonas
22	76.5	150	16	Q52322	Q52322 escherichia
23	76.5	194	10	Q946P3	Q946p3 uncultured
24	76.5	278	10	Q9FXF4	Q9fxf4 arabidopsis
25	76.5	294	10	Q9M2H4	Q9m2h4 arabidopsis
26	76.5	301	16	Q8XUF3	Q8xuf3 ralstonia s
27	76.5	324	10	Q9ZPQ5	Q9zpq5 arabidopsis
28	76.5	363	4	Q9H266	Q9h266 homo sapien
29	76.5	490	16	Q92T24	Q92tz4 rhizobium m
30	76.5	498	16	Q9KN11	Q9kn11 vibrio chol
31	76.5	533	16	Q92BS4	Q92bs4 listeria in
32	76.5	533	16	Q8Y770	Q8y770 listeria mo
33	76.5	543	16	Q9KDL2	Q9kd12 bacillus ha
34	76.5	559	1	Q9UWN9	Q9uwn9 sulfolobus
35	76.5	559	1	Q55088	Q55088 sulfolobus
36	76.5	561	17	P95867	P95867 sulfolobus
37	76.5	562	17	Q9U2L0	Q9uz10 pyrococcus
38	76.5	564	17	Q8U2A4	Q8u2a4 pyrococcus
39	76.5	579	16	Q9KZD3	Q9kzd3 streptomyce
40	76.5	587	16	Q9JZP8	Q9jzp8 neisseria m
41	76.5	587	16	Q9JUT3	Q9jut3 neisseria m
42	76.5	623	16	Q9PAA1	Q9paal xyella fas
43	76.5	642	3	Q9C260	Q9c260 neurospora
44	76.5	778	11	Q99MP4	Q99mp4 mus musculu
45	76.5	791	4	Q9P2E2	Q9p2e2 homo sapien

ALIGNMENTS

RESULT 1

Q8U3Z7	PRELIMINARY;	PRT;	284 AA.
ID	Q8U3Z7		
AC	Q8U3Z7;		
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Co-factor modifying protein.		
GN	PF0302.		
OS	Pyrococcus furiosus.		
OC	Archaea; Euryarchaeota; Thermococci; Thermococcaceae; Thermococcaceae;		
OC	Pyrococcus.		
OX	NCBI_TaxID=2261;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=VCI / DSM 3638 / ATCC 43587 / JCM 8422;		
RA	Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;		
RT	"The complete sequence of the Pyrococcus furiosus genome."		
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AE010154; AAL80426.1; -		
KW	Complete proteome.		
SQ	SEQUENCE 284 AA; 32512 MW; BI9557B14451FFCD CRC64;		

Query Match	88.2%;	Score 30;	DB 17;	Length 284;
Best Local Similarity	62.5%;	Pred. No. 15;		
Matches	5;	Conservative	2;	Mismatches 1; Indels 0; Gaps 0;
Qy	1	EVMXVAEF 8		
Db	221	EIMKIAEF 228		
RESULT 2				
Q97F65				
ID	Q97F65	PRELIMINARY;	PRT;	195 AA.
AC	Q97F65;			
DT	01-OCT-2001 (TREMBLrel. 18, Created)			
DT	01-OCT-2001 (TREMBLrel. 18, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			

```

DE Thymidine kinase.
GN CAC2887.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tabsov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007786; AAK80830.1; -.
DR InterPro; IPR000345; Cytochrome_b.
DR Pfam; PF00265; TK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00603; TK_CELLULAR_TYPE; UNKNOWN_1.
KW Kinase; Complete proteome.
SQ SEQUENCE 195 AA; 22303 MW; 6EDB551986B203BA CRC64;

Query Match 85.3%; Score 29; DB 16; Length 195;
Best Local Similarity 62.5%; Pred. No. 19;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8
   I : I : I I I
Db 130 ELMAIAEF 137

RESULT 3
Q800A9 PRELIMINARY; PRT; 262 AA.
AC Q800A9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 2-dehydro-3-deoxyphosphoheptamate aldolase (EC 4.1.2.15).
GN PF1690.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE010267; AAL81814.1; -.
KW Lyase; Complete proteome.
SQ SEQUENCE 262 AA; 29225 MW; ED84D28FD5DEADC2 CRC64;

Query Match 82.4%; Score 28; DB 17; Length 262;
Best Local Similarity 62.5%; Pred. No. 47;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8
   : : I I I I I
Db 38 QIMKVAEF 45

RESULT 4
Q82PFO PRELIMINARY; PRT; 594 AA.
AC Q82PFO;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

```

```

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative alpha amylase.
GN STM1560.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Willson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
DR EMBL; AE008768; AAL20478.1; -.
DR InterPro; IPR00461; Alpha_amylase.
DR InterPro; IPR004193; Isoamylase_N.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF02922; isoamylase_N; 1.
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 594 AA; 65774 MW; B4C685D611C7E38B CRC64;

Query Match 82.4%; Score 28; DB 16; Length 594;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8
   I I I I I
Db 148 EVMPVAQF 155

RESULT 5
Q82724 PRELIMINARY; PRT; 594 AA.
AC Q82724;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative hydrolase.
GN STY1503.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RL Nature 413:848-852(2001).
DR EMBL; AL627270; CAD01762.1; -.
DR InterPro; IPR00461; Alpha_amylase.
DR InterPro; IPR004193; Isoamylase_N.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF02922; isoamylase_N; 1.
KW Hydrolase; Hypothetical protein; Complete proteome.
SQ SEQUENCE 594 AA; 65816 MW; 4D128BD3D2D3980E CRC64;

Query Match 82.4%; Score 28; DB 16; Length 594;

```

Best Local Similarity 75.0%; Pred. No. 1e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
||| |||

Db 148 EVMVAQF 155

# RESULT 6

Q8XT78 PRELIMINARY; PRT; 622 AA.  
AC Q8XT78: 01-MAR-2002 (TRENBLrel. 20, Created)  
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE Putative MALTOOLIGOSYL trehalose TREHALOHYDROLASE protein.  
GN RSP0237 OR RS05186.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OG Plasmid megaplasmid.  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GM11000;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
Chandler M., Cholsne N., Claudel-Renard C., Cunac S., Demange N.,  
Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,  
Weissenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."  
RL Nature 415:497-502(2002).  
DR EMBL: AL646077; CAD17388.1; -;  
DR InterPro; IPR000461; Alpha-amylase.  
DR InterPro; IPR004193; Isoamylase-N.  
DR Pfam; PF00128; alpha-amylase; 1.  
DR Pfam; PF02922; isoamylase-N; 1.  
KW Plasmid; Complete proteome.  
SQ SEQUENCE 622 AA; 68369 MW; 9F639FA9DFB9BB21 CRC64;

Query Match 82.4%; Score 28; DB 16; Length 622;  
Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
|:| ||||

Db 157 ELMVAEF 164

# RESULT 7

Q94IL7 PRELIMINARY; PRT; 776 AA.  
AC Q94IL7: 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE Putative high-affinity potassium uptake transporter.  
GN KUP1.  
OS Populus tremula x Populus tremuloides.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eudicots I; Malpighiales; Salicaceae; Populus.  
OX NCBI\_TaxID=47664;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Langer K., Ache P., Fromm J., Hedrich R.;  
RT "PTKUP contributes to wood formation."  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ299422; CAC93168.1; -;  
DR InterPro; IPR003855; K+-transprtr.  
DR Pfam; PF02705; K\_trans; 1.

DR TIGRFAMS; TIGR00794; kup; 1.  
SQ SEQUENCE 776 AA; 87303 MW; BEC03D57ED0869BC CRC64;

Query Match 82.4%; Score 28; DB 10; Length 776;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VMXVAEF 8  
|| ||||

Db 604 VMSVAEF 610

# RESULT 8

Q8YVT6 PRELIMINARY; PRT; 1256 AA.  
AC Q8YVT6: 01-MAR-2002 (TRENBLrel. 20, Created)  
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
DE Hypothetical protein All1888.  
GN All1888.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
Watanabe A., Iriuchi M., Ishikawa A., Kawashima K., Kimura T.,  
Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,  
Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
cyanobacterium Anabaena sp. strain PCC 7120."  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003587; BAB73587.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1256 AA; 147253 MW; DDFPB5E19241BAC6 CRC64;

Query Match 82.4%; Score 28; DB 16; Length 1256;  
Best Local Similarity 75.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
||| |||

Db 74 EVMVAEF 81

# RESULT 9

Q9IG40 PRELIMINARY; PRT; 3414 AA.  
AC Q9IG40: 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE Polyprotein precursor.  
OS Langat virus (strain TP21).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Flavivirus.  
OX NCBI\_TaxID=31638;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TP21;  
RX MEDLINE=20192178; PubMed=10725214;  
RA Campbell M.S., Pietnev A.G.;  
RT "Infectious cDNA clones of Langat tick-borne flavivirus that differ  
from their parent in peripheral neurovirulence."  
RL Virology 269:225-237(2000).  
DR EMBL; AF253419; AAF75259.1; -;  
DR HSP; P14336; 1SVB.  
DR MEROPS; S07.001; -;  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR001122; Flavi\_capsidc.

DR InterPro: IPR000336; Flavi\_glycoprote. 82.4%; Score 28; DB 12; Length 3414;  
 DR InterPro: IPR001850; Flavi\_helicase. 85.7%; Pred. No. 5.6e+02; Mismatches 0; Indels 0; Gaps 0;  
 DR InterPro: IPR000069; Flavi\_M. 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 DR InterPro: IPR001157; Flavi\_NS1. 2 VMXVAEF 8  
 DR InterPro: IPR000752; Flavi\_NS2A. 111111  
 DR InterPro: IPR000487; Flavi\_NS2B. 929 VMTVAEF 935  
 DR InterPro: IPR000404; Flavi\_NS4A. PRELIMINARY; PRT; 3414 AA.  
 DR InterPro: IPR001528; Flavi\_NS4B. 01-OCT-2000 (TRENBLrel. 15, Created)  
 DR InterPro: IPR000208; Flavi\_NS5. 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DR InterPro: IPR002535; Flavi\_propep. 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DR InterPro: IPR002877; FtsJ. Polypeptide precursor.  
 DR InterPro: IPR001650; Helicase\_C. Langkat virus.  
 DR Pfam: PF01003; Flavi\_capsid; 1. Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 DR Pfam: PF00869; Flavi\_glycoprote; 1. Flavivirus.  
 DR Pfam: PF02832; Flavi\_glycopC; 1. NCBI\_TaxID=11085;  
 DR Pfam: PF00949; Flavi\_helicase; 1. SEQUENCE FROM N.A.  
 DR Pfam: PF01004; Flavi\_M; 1. STRAIN-ATTENUATED STRAIN E5;  
 DR Pfam: PF00948; Flavi\_NS1; 1. MEDLINE=20192178; PubMed=10725214;  
 DR Pfam: PF01005; Flavi\_NS2A; 1. Campbell M.S.; Pletnev A.G.;  
 DR Pfam: PF01002; Flavi\_NS2B; 1. "Infectious cDNA clones of Langkat tick-borne flavivirus that differ  
 DR Pfam: PF01350; Flavi\_NS4A; 1. from their parent in peripheral neurovirulence."  
 DR Pfam: PF01349; Flavi\_NS4B; 1.  
 DR Pfam: PF00972; Flavi\_NS5; 1.  
 DR Pfam: PF01570; Flavi\_propep; 1.  
 DR Pfam: PF01728; FtsJ; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR ProDom: PD001496; Flavi\_NS1; 1.  
 DR ProDom: PD001556; Flavi\_glycoprote; 1.  
 DR SMART; SM00490; HELICC; 1.  
 KW Signal.  
 FT CHAIN 97 116 POTENTIAL.  
 FT CHAIN 1 96 CAPSID PROTEIN.  
 FT CHAIN 117 280 MEMBRANE-ASSOCIATED GLYCOPROTEIN  
 FT CHAIN 281 779 PRECURSOR  
 FT CHAIN 780 1128 ENVELOPE MEMBRANE-ASSOCIATED  
 FT CHAIN 1129 1358 GLYCOPROTEIN.  
 FT CHAIN 1359 1489 NONSTRUCTURAL PROTEIN NS1.  
 FT CHAIN 1490 2110 NONSTRUCTURAL PROTEIN NS2A.  
 FT CHAIN 2111 2259 NONSTRUCTURAL PROTEIN NS2B.  
 FT CHAIN 2260 2511 NONSTRUCTURAL PROTEIN NS3.  
 FT CHAIN 2512 3414 NONSTRUCTURAL PROTEIN NS4A.  
 FT CHAIN 3414 AA; 378135 MW; 8DE86A46A0E8F7E3 CRC64; NONSTRUCTURAL PROTEIN NS4B.  
 SQ SEQUENCE 3414 AA; 378135 MW; 8DE86A46A0E8F7E3 CRC64; NONSTRUCTURAL PROTEIN NS5.

Query Match 82.4%; Score 28; DB 12; Length 3414;  
 Best Local Similarity 85.7%; Pred. No. 5.6e+02; Mismatches 0; Indels 0; Gaps 0;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 VMXVAEF 8  
 111111  
 Db 929 VMTVAEF 935  
 PRELIMINARY; PRT; 3414 AA.  
 RESULT 10  
 Q9IG39  
 ID O9IG39 PRELIMINARY; PRT; 3414 AA.  
 AC O9IG39;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Polypeptide precursor.  
 OS Langkat virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus.  
 OX NCBI\_TaxID=11085;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATTENUATED STRAIN E5;  
 RX MEDLINE=20192178; PubMed=10725214;  
 RA Campbell M.S.; Pletnev A.G.;  
 RT "Infectious cDNA clones of Langkat tick-borne flavivirus that differ  
 from their parent in peripheral neurovirulence.";

RL Virology 269:225-237(2000).  
 DR EMBL; AF253420; AAF75260.1;  
 DR HSSP; P14336; 1SVB.  
 DR MEROPS; S07\_001;  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR001122; Flavi\_capsidC.  
 DR InterPro: IPR000336; Flavi\_glycoprote.  
 DR InterPro: IPR001850; Flavi\_helicase.  
 DR InterPro: IPR000069; Flavi\_M.  
 DR InterPro: IPR001157; Flavi\_NS1.  
 DR InterPro: IPR000752; Flavi\_NS2A.  
 DR InterPro: IPR000487; Flavi\_NS2B.  
 DR InterPro: IPR000404; Flavi\_NS4A.  
 DR InterPro: IPR001528; Flavi\_NS4B.  
 DR InterPro: IPR000208; Flavi\_NS5.  
 DR InterPro: IPR002535; Flavi\_propep.  
 DR InterPro: IPR002877; FtsJ.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR Pfam: PF01003; Flavi\_capsid; 1.  
 DR Pfam: PF00869; Flavi\_glycoprote; 1.  
 DR Pfam: PF02832; Flavi\_glycopC; 1.  
 DR Pfam: PF00949; Flavi\_helicase; 1.  
 DR Pfam: PF01004; Flavi\_M; 1.  
 DR Pfam: PF00948; Flavi\_NS1; 1.  
 DR Pfam: PF01005; Flavi\_NS2A; 1.  
 DR Pfam: PF01002; Flavi\_NS2B; 1.  
 DR Pfam: PF01350; Flavi\_NS4A; 1.  
 DR Pfam: PF01349; Flavi\_NS4B; 1.  
 DR Pfam: PF00972; Flavi\_NS5; 1.  
 DR Pfam: PF01570; Flavi\_propep; 1.  
 DR Pfam: PF01728; FtsJ; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR ProDom: PD001496; Flavi\_NS1; 1.  
 DR ProDom: PD001556; Flavi\_glycoprote; 1.  
 DR SMART; SM00490; HELICC; 1.  
 KW Signal.  
 FT CHAIN 97 116 POTENTIAL.  
 FT CHAIN 1 96 CAPSID PROTEIN.  
 FT CHAIN 117 280 MEMBRANE-ASSOCIATED GLYCOPROTEIN  
 FT CHAIN 281 779 PRECURSOR  
 FT CHAIN 780 1128 ENVELOPE MEMBRANE-ASSOCIATED  
 FT CHAIN 1129 1358 GLYCOPROTEIN.  
 FT CHAIN 1359 1489 NONSTRUCTURAL PROTEIN NS1.  
 FT CHAIN 1490 2110 NONSTRUCTURAL PROTEIN NS2A.  
 FT CHAIN 2111 2259 NONSTRUCTURAL PROTEIN NS2B.  
 FT CHAIN 2260 2511 NONSTRUCTURAL PROTEIN NS3.  
 FT CHAIN 2512 3414 NONSTRUCTURAL PROTEIN NS4A.  
 FT CHAIN 3414 AA; 378014 MW; BF8F8ACEAB96D534 CRC64; NONSTRUCTURAL PROTEIN NS4B.  
 SQ SEQUENCE 3414 AA; 378014 MW; BF8F8ACEAB96D534 CRC64; NONSTRUCTURAL PROTEIN NS5.

Query Match 82.4%; Score 28; DB 12; Length 3414;  
 Best Local Similarity 85.7%; Pred. No. 5.6e+02; Mismatches 0; Indels 0; Gaps 0;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 VMXVAEF 8  
 111111  
 Db 929 VMTVAEF 935  
 PRELIMINARY; PRT; 184 AA.  
 RESULT 11  
 O26364  
 ID O26364 PRELIMINARY; PRT; 184 AA.  
 AC O26364;  
 DT 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE DNA-dependent RNA polymerase, subunit E'.  
 GN MTH264.  
 OS Methanobacterium thermoautotrophicum.  
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
 OC Methanobacteriaceae; Methanothermobacter.  
 OX NCBI\_TaxID=187420;

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-DELTA H;
RX  MEDLINE=98037514; PubMed=9371463;
RA  Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
RA  Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA  Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA  Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA  Jiwni N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA  McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA  Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT  "Complete genome sequence of Methanobacterium thermoautotrophicum
RT  deltaH: functional analysis and comparative genomics.";
RL  J. Bacteriol. 179:7135-7155(1997).
DR  EMBL: AE000812; AAB84770.1; -;
DR  HSSP: P05055; 1SRO.
DR  InterPro: IPR004519; RpoE.
DR  InterPro: IPR003029; S1.
DR  Pfam: PF00575; S1; 1.
DR  SMART: SM00316; S1; 1.
DR  TIGRFAMS: TIGR00448; rpoE; 1.
KW  Complete proteome.
SQ  SEQUENCE 184 AA; 20695 MW; 37AAB7B39D631813 CRC64;

Query Match 79.4%; Score 27; DB 17; Length 184;
Best Local Similarity 62.5%; Pred. No. 61;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8
DB 88 EVIEIAEF 95

RESULT 12
QHLL69
ID Q9HL69 PRELIMINARY; PRT; 338 AA.
AC Q9HL69;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein Ta0362.
GN Ta0362.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Newes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum.";
RL Nature 407:508-513(2000).
DR EMBL: AL445064; CAC11506.1; -;
DR InterPro: IPR001005; Myb DNA binding.
DR PROSITE: PS00037; MYB.1; UNKNOWN.1.
KW Hypothetical protein, Complete proteome.
SQ SEQUENCE 338 AA; 39348 MW; D86C955529E9A43E CRC64;

Query Match 79.4%; Score 27; DB 17; Length 338;
Best Local Similarity 62.5%; Pred. No. 1; Indels 0; Gaps 0;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8
DB 329 EIMPLAEF 336

RESULT 13
OS8774
ID O58774 PRELIMINARY; PRT; 598 AA.

```

```

AC O58774;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein PH1023.
GN PH1023.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kwarabavasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo H., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL: AP000004; BAA30120.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 598 AA; 70188 MW; 3E1ACD0F8EDC7661 CRC64;

Query Match 79.4%; Score 27; DB 17; Length 598;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8
DB 280 EIVNVAEF 287

RESULT 14
Q92U63
ID Q92U63 PRELIMINARY; PRT; 601 AA.
AC Q92U63;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative 1,4-alpha-glucan branching enzyme protein (EC
DE 2.4.1.18).
GN GLGB2 OR RB1276 OR SMB21447.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoeiter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL: AL603646; CAC49676.1; -;
DR InterPro: IPR000461; Alpha-amylase.
DR Pfam: PF00128; isoamylase_N.
DR Pfam: PF02922; isoamylase_N.1.
DR Transferrase; Glycosyltransferase; Plasmid; Hypothetical protein;
KW Complete proteome.
SQ SEQUENCE 601 AA; 66954 MW; 024DA322A7B72C2E CRC64;

Query Match 79.4%; Score 27; DB 16; Length 601;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 EVMXVAEF 8  
 I: I I I I  
 Db 151 EIMPVAQF 158

## RESULT 15

Q9FE38 PRELIMINARY; PRT; 775 AA.  
 AC Q9FE38;  
 DT 01-MAR-2001 (TREMELrel. 16, Created)  
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)  
 DE Tiny root hair 1 protein.  
 GN TH1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rigas S., Debrosses G., Haralampidis K., Vicente-Agullo F.,  
 RA Feldmann K.A., Grabov A., Dolan L., Hatzopoulos P.;  
 RT "A Potassium Transporter Required for Tip Growth in Arabidopsis.";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ296156; CAC16138.1;  
 DR EMBL; AJ296155; CAC16137.1;  
 DR InterPro; IPR003855; K+-transprtr.  
 DR Pfam; PF02705; K\_trans; 1.  
 DR TIGRFAMs; TIGR00794; kup; 1.  
 SQ SEQUENCE 775 AA; 86842 MW; B0C55068B48E8180 CRC64;

Query Match 79.4%; Score 27; DB 10; Length 775;  
 Best Local Similarity 71.4%; Pred. No. 2.4e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VNXVAEF 8  
 I: I I I I  
 Db 602 VNSIAEF 608

Search completed: April 23, 2003, 12:38:41  
 Job time : 36.2069 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:36:34 ; Search time 17.1034 Seconds  
(without alignments)  
37.480 Million cell updates/sec

Title: US-09-723-722A-81

Perfect score: 34

Sequence: 1 EVMXVAEF 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	79.4	68	10	US-09-393-634-80
2	27	79.4	347	9	US-09-738-626-6922
3	26	76.5	282	10	US-09-925-300-1694
4	26	76.5	559	10	US-09-908-855-15
5	25	73.5	27	10	US-09-864-761-34137
6	25	73.5	54	9	US-09-158-722-18
7	25	73.5	152	10	US-09-741-669-445
8	25	73.5	237	9	US-09-738-626-4458
9	25	73.5	280	9	US-09-738-626-4505
10	25	73.5	479	8	US-08-910-386A-24
11	25	73.5	485	9	US-09-918-543-4
12	25	73.5	485	9	US-09-795-211-2
13	25	73.5	485	10	US-09-769-864-2
14	25	73.5	485	10	US-09-769-864-8
15	25	73.5	485	10	US-09-854-346-4
16	25	73.5	485	10	US-09-902-188A-2
17	25	73.5	499	9	US-10-108-605-283
18	24	70.6	57	10	US-09-879-957-209
19	24	70.6	145	9	US-10-013-379-11

20	70.6	171	10	US-09-815-242-13274	Sequence 13274, A
21	70.6	171	10	US-09-815-242-13532	Sequence 13532, A
22	70.6	335	10	US-09-815-242-13346	Sequence 13346, A
23	70.6	387	9	US-09-895-913A-120	Sequence 120, App
24	70.6	409	9	US-10-081-816-14	Sequence 14, Appl
25	70.6	1084	9	US-10-024-623-23	Sequence 23, Appl
26	70.6	1095	9	US-10-024-623-26	Sequence 26, Appl
27	70.6	1095	9	US-10-024-623-36	Sequence 36, Appl
28	70.6	1854	9	US-10-029-413A-2	Sequence 2, Appl1
29	70.6	1873	9	US-10-029-413A-12	Sequence 12, Appl
30	70.6	1873	9	US-10-029-413A-14	Sequence 14, Appl
31	67.6	54	9	US-09-158-722-46	Sequence 46, Appl
32	67.6	54	9	US-09-158-722-47	Sequence 47, Appl
33	67.6	90	9	US-09-510-332-40	Sequence 40, Appl
34	67.6	90	10	US-09-393-634-76	Sequence 76, Appl
35	67.6	150	10	US-09-815-242-4899	Sequence 4899, Ap
36	67.6	160	9	US-09-931-457A-7	Sequence 7, Appl1
37	67.6	167	9	US-10-102-627-65	Sequence 65, Appl
38	67.6	205	9	US-10-042-141-103	Sequence 103, App
39	67.6	205	10	US-09-726-643-103	Sequence 103, App
40	67.6	290	10	US-09-815-242-13182	Sequence 13182, A
41	67.6	310	9	US-09-939-833-7	Sequence 7, Appli
42	67.6	310	10	US-09-939-754-7	Sequence 7, Appli
43	67.6	310	10	US-09-939-832-7	Sequence 7, Appli
44	67.6	314	9	US-09-510-332-44	Sequence 44, Appl
45	67.6	314	9	US-09-510-332-71	Sequence 71, Appl

#### ALIGNMENTS

#### RESULT 1

US-09-393-634-80  
; Sequence 80, Application US/09393634  
; Patent No. US20020051997A1  
; GENERAL INFORMATION:  
; APPLICANT: Zuker, Charles S.  
; APPLICANT: Adler, Jon Elliot  
; APPLICANT: Ryba, Nick  
; APPLICANT: Mueller, Ken  
; APPLICANT: Hoon, Mark  
; APPLICANT: The Regents of the University of California  
; APPLICANT: The Government of the United States of America  
; APPLICANT: as represented by the Secretary of the  
; APPLICANT: Department of Health and Human Services  
; TITLE OF INVENTION: SF, a No. US20020051997A1el Family of Taste Receptors  
; FILE REFERENCE: 02307E-09800005  
; CURRENT APPLICATION NUMBER: US/09/393, 634  
; CURRENT FILING DATE: 1999-09-10  
; NUMBER OF SEQ ID NOS: 92  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 80  
; LENGTH: 68  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human GR24  
US-09-393-634-80

Query Match 79.4%; Score 27; DB 10; Length 68;  
Best Local Similarity 71.4%; Pred. No. 9.6;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VMXVAEF 8

Db 11 IMAVAEF 17

#### RESULT 2

US-09-738-626-6922  
; Sequence 6922, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIALI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAOKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OKAZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 6922  
LENGTH: 347  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6922

Query Match 79.4%; Score 27; DB 9; Length 347;  
Best Local Similarity 62.5%; Pred. No. 53;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
|::|  
DB 78 EILVAEF 85

RESULT 3  
US-09-925-300-1694  
Sequence 1694, Application US/09925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Rosen,  
APPLICANT: Steve Ruben  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/09/925,300  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn ver. 2.0  
SEQ ID NO 1694  
LENGTH: 282  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (20)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-300-1694

Query Match 76.5%; Score 26; DB 10; Length 282;  
Best Local Similarity 75.0%; Pred. No. 74;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
|||  
DB 162 EVKDVAEF 169

RESULT 4  
US-09-908-855-15  
Sequence 15, Application US/09908855  
Patent No. US20020115220A1  
GENERAL INFORMATION:  
APPLICANT: KONDO, KEIJI  
APPLICANT: MIURA, YUTAKA  
TITLE OF INVENTION: YEAST VECTOR AND METHOD OF PRODUCING PROTEINS USING THE  
TITLE OF INVENTION: SAME  
FILE REFERENCE: 049441/0118  
CURRENT APPLICATION NUMBER: US/09/908,855  
CURRENT FILING DATE: 2001-07-20  
PRIOR APPLICATION NUMBER: 09/242,690  
PRIOR FILING DATE: 1999-02-23  
PRIOR APPLICATION NUMBER: JP 8/241062  
PRIOR FILING DATE: 1996-08-23  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: PatentIn ver. 2.1  
SEQ ID NO 15  
LENGTH: 559  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein  
US-09-908-855-15

Query Match 76.5%; Score 26; DB 10; Length 559;  
Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
|::|  
DB 138 EIMPIAQF 145

RESULT 5  
US-09-864-761-34137  
Sequence 34137, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 34137  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AP000118.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.8  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.6  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 15  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 18  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 34  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 25  
; OTHER INFORMATION: EST\_HUMAN HIT: AV703969.1, EVALUATE 1.20e-02  
; OTHER INFORMATION: SWISSPROT HIT: Q24186, EVALUATE 1.00e-04  
US-09-864-761-34137

Query Match 73.5%; Score 25; DB 10; Length 27;  
Best Local Similarity 57.1%; Pred. No. 11;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVMXVAE 7  
| : | : | : |  
Db 12 EIMTIAE 18

RESULT 6  
US-09-158-722-18  
; Sequence 18, Application US/09158722  
; Publication No. US20030013848A1  
; GENERAL INFORMATION:  
; APPLICANT: Lemke Ph.D. et al., Greg E.  
; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/158,722  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/456,647  
; FILING DATE: 02-JUN-1995  
; APPLICATION NUMBER: US 08/237,401  
; FILING DATE: 02-MAY-1994  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/884,486  
; FILING DATE: 15-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wetherell Ph.D., John R.  
; REGISTRATION NUMBER: 31,678  
; REFERENCE/DOCKET NUMBER: 07251/007002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 678-5070  
; TELEFAX: (619) 678-5099  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 54 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-158-722-18

Query Match 73.5%; Score 25; DB 9; Length 54;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVMXVAEF 8  
| : | : | : |  
Db 8 DVMKIADF 15

RESULT 7  
US-09-741-669-445  
; Sequence 445, Application US/09741669  
; Patent No. US20020022718A1  
; GENERAL INFORMATION:  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; TITLE OF INVENTION: Genes identified as required for  
; FILE OF INVENTION: proliferation of E. coli  
; FILE REFERENCE: ELIWA.009A  
; CURRENT APPLICATION NUMBER: US/09/741,669  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 60/173005  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 445  
; LENGTH: 152  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-669-445

Query Match 73.5%; Score 25; DB 10; Length 152;  
Best Local Similarity 62.5%; Pred. No. 68;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVMXVAEF 8  
| : | : | : |  
Db 107 EVMLVGQF 114

RESULT 8  
US-09-738-626-4458  
; Sequence 4458, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: OCHIALI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4458  
; LENGTH: 237  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4458

Query Match 73.5%; Score 25; DB 9; Length 237;  
Best Local Similarity 37.5%; Pred. No. 1.1e+02;  
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
Db 73 QILSIAEF 80

RESULT 9  
US-09-738-626-4505  
; Sequence 4505, Application US/09738626  
; Publication No. US20020197603A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OKAZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4505  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4505

Query Match 73.5%; Score 25; DB 9; Length 280;  
Best Local Similarity 62.5%; Pred. No. 1.3e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
Db 5 DVVVVAEF 12

RESULT 10  
US-08-910-386A-24  
; Sequence 24, Application US/08910386A  
; Patent No. US20020092041A1

; GENERAL INFORMATION:  
; APPLICANT: Ronald, Pamela C.  
; APPLICANT: Wang, Guo-Liang  
; APPLICANT: Song, Wen-Yuang  
; APPLICANT: Hulbert, Scot  
; APPLICANT: Richter, Todd  
; TITLE OF INVENTION: Procedures and Materials for Conferring  
; TITLE OF INVENTION: Disease Resistance in Plants  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/910,386A  
; FILING DATE: 13-AUG-1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 023070-0589500S  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 479 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-910-386A-24

Query Match 73.5%; Score 25; DB 8; Length 479;  
Best Local Similarity 62.5%; Pred. No. 2.3e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
Db 163 EVMALLEF 170

RESULT 11  
US-09-918-543-4  
; Sequence 4, Application US/09918543  
; Patent No. US20020155574A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20020155574Alozymes A/S  
; APPLICANT: Thisted, Thomas  
; APPLICANT: Kjaerulff, Soren  
; APPLICANT: Andersen, Carsten  
; APPLICANT: Fuglsang, Claus Crone  
; TITLE OF INVENTION: Alpha-amylase mutants with altered properties  
; FILE REFERENCE: 10062.200-US  
; CURRENT APPLICATION NUMBER: US/09/918,543  
; CURRENT FILING DATE: 2001-07-31  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 485  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
US-09-918-543-4

Query Match 73.5%; Score 25; DB 9; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.3e+02;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
I: |||||

Db 260 EMFAVAEF 267

RESULT 12

US-09-795-211-2

; Sequence 2, Application US/09795211

; Publication No. US20020183226A1

; GENERAL INFORMATION:

; APPLICANT: The Procter & Gamble Company

; APPLICANT: Kasturi, Chandrika

; APPLICANT: Wandstrat, Mark E.

; APPLICANT: Song, Brian X.

; APPLICANT: Nissen, Torben L.

; APPLICANT: Kjaerulff, Soren

; TITLE OF INVENTION: LIQUID DETERGENT COMPOSITION EXHIBITING ENHANCED a-AMYLASE ENZYME ACTIVITY

; FILE REFERENCE: 5368.200-US

; CURRENT APPLICATION NUMBER: US/09/795,211

; PRIOR FILING DATE: 2001-02-28

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 485

; TYPE: PRT

; ORGANISM: alakaliphilicbacillus

US-09-795-211-2

Query Match 73.5%; Score 25; DB 9; Length 485;

Best Local Similarity 62.5%; Pred. No. 2.3e+02;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
I: |||||

Db 260 EMFAVAEF 267

RESULT 13

US-09-769-864-2

; Sequence 2, Application US/09769864

; Patent No. US20010039253A1

; GENERAL INFORMATION:

; APPLICANT: Borchert, Torben V.

; APPLICANT: Svendsen, Allan

; APPLICANT: Andersen, Carsten

; APPLICANT: Nielsen, Bjarne

; APPLICANT: Nissen, Torben L.

; APPLICANT: Kjaerulff, Soren

; TITLE OF INVENTION: Alpha-Amulase Mutants

; FILE REFERENCE: 5368.200-US

; CURRENT APPLICATION NUMBER: US/09/769,864

; PRIOR FILING DATE: 2001-01-25

; PRIOR APPLICATION NUMBER: 09/183,412

; NUMBER OF SEQ ID NOS: 58

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 485

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-769-864-2

Query Match 73.5%; Score 25; DB 10; Length 485;

Best Local Similarity 62.5%; Pred. No. 2.3e+02;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
I: |||||

Db 260 EMFAVAEF 267

RESULT 14

US-09-769-864-8

; Sequence 8, Application US/09769864

; Patent No. US20010039253A1

; GENERAL INFORMATION:

; APPLICANT: Borchert, Torben V.

; APPLICANT: Svendsen, Allan

; APPLICANT: Andersen, Carsten

; APPLICANT: Nielsen, Bjarne

; APPLICANT: Nissen, Torben L.

; APPLICANT: Kjaerulff, Soren

; TITLE OF INVENTION: Alpha-Amulase Mutants

; FILE REFERENCE: 5368.200-US

; CURRENT APPLICATION NUMBER: US/09/769,864

; PRIOR FILING DATE: 2001-01-25

; PRIOR APPLICATION NUMBER: 09/183,412

; PRIOR FILING DATE: 1998-10-30

; NUMBER OF SEQ ID NOS: 58

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 8

; LENGTH: 485

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-769-864-8

Query Match 73.5%; Score 25; DB 10; Length 485;

Best Local Similarity 62.5%; Pred. No. 2.3e+02;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
I: |||||

Db 260 EMFAVAEF 267

RESULT 15

US-09-854-346-4

; Sequence 4, Application US/09854346

; Patent No. US20020068352A1

; GENERAL INFORMATION:

; APPLICANT: Svendsen, Allan

; APPLICANT: Jorgensen, Christel Thea

; APPLICANT: Nielsen, Bjarne Ronfeldt

; TITLE OF INVENTION: Alpha-amylase variants with altered 1,6 activity

; FILE REFERENCE: 6140.200-US

; CURRENT APPLICATION NUMBER: US/09/854,346

; CURRENT FILING DATE: 2001-05-11

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 4

; LENGTH: 485

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-854-346-4

Query Match 73.5%; Score 25; DB 10; Length 485;

Best Local Similarity 62.5%; Pred. No. 2.3e+02;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
I: |||||

Db 260 EMFAVAEF 267

Search completed: April 23, 2003, 12:49:51

Job time : 18.1034 secs



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM. protein - protein search, using sw model

Run on: April 23, 2003, 12:35:24 ; Search time 12.9655 Seconds  
(without alignments)  
18.155 Million cell updates/sec

Title: US-09-723-722A-81  
Perfect score: 34  
Sequence: 1 EVMXVAEF 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTC05\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	76.5	559	US-09-242-690A-15	Sequence 15, Appl
2	26	76.5	559	US-09-298-924-6	Sequence 6, Appl
3	25	73.5	54	US-08-456-647B-18	Sequence 18, Appl
4	25	73.5	54	US-08-237-401A-18	Sequence 18, Appl
5	25	73.5	405	US-09-291-023A-20	Sequence 20, Appl
6	25	73.5	485	US-08-446-803-2	Sequence 2, Appl
7	25	73.5	485	US-08-861-837-2	Sequence 2, Appl
8	25	73.5	485	US-08-600-656-2	Sequence 2, Appl
9	25	73.5	485	US-09-170-670-2	Sequence 2, Appl
10	25	73.5	485	US-09-170-670-8	Sequence 2, Appl
11	25	73.5	485	US-09-193-068-2	Sequence 2, Appl
12	25	73.5	485	US-09-193-068-8	Sequence 2, Appl
13	25	73.5	485	US-09-183-412-2	Sequence 2, Appl
14	25	73.5	485	US-09-183-412-8	Sequence 2, Appl
15	25	73.5	485	US-09-264-097-5	Sequence 5, Appl
16	25	73.5	485	US-09-354-191A-2	Sequence 2, Appl
17	25	73.5	485	US-09-290-734-2	Sequence 2, Appl
18	25	73.5	485	US-09-290-734-8	Sequence 8, Appl
19	25	73.5	513	US-08-459-346-19	Sequence 19, Appl
20	25	73.5	513	US-07-989-847-8	Sequence 8, Appl
21	25	73.5	513	US-08-889-419-19	Sequence 19, Appl
22	25	73.5	513	US-08-469-411-8	Sequence 8, Appl
23	25	73.5	513	US-08-402-542-19	Sequence 19, Appl
24	25	73.5	513	PCT-US93-07189-19	Sequence 19, Appl
25	25	73.5	513	5187076-6	Patent No. 5187076
26	25	73.5	556	US-08-505-377-1	Sequence 1, Appl
27	25	73.5	556	US-08-798-269-1	Sequence 1, Appl

28 25 73.5 556 4 US-09-055-210-1 Sequence 1, Appl  
29 25 73.5 556 4 US-09-298-924-8 Sequence 8, Appl  
30 25 73.5 816 1 US-07-640-029-1 Sequence 1, Appl  
31 25 73.5 817 1 US-07-640-029-2 Sequence 2, Appl  
32 24 70.6 57 4 US-08-630-915A-209 Sequence 209, App  
33 24 70.6 161 4 US-08-858-207A-284 Sequence 284, App  
34 24 70.6 329 1 US-08-230-047-7 Sequence 7, Appl  
35 24 70.6 341 1 US-08-314-309A-19 Sequence 19, Appl  
36 24 70.6 347 4 US-09-739-455-5 Sequence 5, Appl  
37 24 70.6 347 4 US-09-739-455-15 Sequence 15, Appl  
38 24 70.6 524 3 US-08-557-210A-3 Sequence 3, Appl  
39 24 70.6 539 3 US-08-557-210A-4 Sequence 4, Appl  
40 24 70.6 539 3 US-08-557-210A-5 Sequence 5, Appl  
41 24 70.6 620 4 US-08-637-670-40 Sequence 40, Appl  
42 24 70.6 691 5 PCT-US91-08442-2 Sequence 2, Appl  
43 24 70.6 758 1 US-07-756-250-16 Sequence 16, Appl  
44 24 70.6 927 4 US-09-134-001C-4831 Sequence 4831, Ap  
45 24 70.6 972 3 US-08-335-844A-34 Sequence 24, Appl

## ALIGNMENTS

### RESULT 1

US-09-242-690A-15  
; Sequence 15, Application US/09242690A  
; Patent No. 6284534

GENERAL INFORMATION:  
; APPLICANT: KONDO, KEIJI

; APPLICANT: MIURA, YUTAKA  
; TITLE OF INVENTION: YEAST VECTOR AND METHOD OF PRODUCING PROTEINS USING THE

; TITLE OF INVENTION: SAME  
; FILE REFERENCE: 049441/0118

; CURRENT APPLICATION NUMBER: US/09/242,690A  
; PRIOR FILING DATE: 1999-02-23

; PRIOR APPLICATION NUMBER: PCT/JP97/02924  
; PRIOR FILING DATE: 1997-08-22

; PRIOR APPLICATION NUMBER: JP 8/241062  
; PRIOR FILING DATE: 1996-08-23

; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 15  
; LENGTH: 559

; TYPE: PRT  
; ORGANISM: Artificial Sequence

; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein

US-09-242-690A-15  
Query Match 76.5%; Score 26; DB 4; Length 559;  
Best Local Similarity 50.0%; Pred No. 1.6e+02;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
Db 138 EIMPIAQF 145  
1:1:1:1

### RESULT 2

US-09-298-924-6  
; Sequence 6, Application US/09298924  
; Patent No. 6391595

GENERAL INFORMATION:  
; APPLICANT: KATO, Masaru

; APPLICANT: MIURA, Yutaka  
; APPLICANT: KETTOKU, Masako

; APPLICANT: IWAMATSU, Akihiro  
; APPLICANT: KOBAYASHI, Kazuo

; APPLICANT: KOMEDA, Toshihiro  
; APPLICANT: NOVEL TRANSFERASE AND AMYLASE, PROCESS

; APPLICANT: FOR PRODUCING THE ENZYMES, USE THEREOF, AND GENE CODING  
; APPLICANT: FOR THE SAME

; NUMBER OF SEQUENCES: 58

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Foley & Lardner  
;; STREET: 3000 K Street, N.W., Suite 500  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: US  
;; ZIP: 20007-5109  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/298,924  
;; FILING DATE: 26-Apr-1999  
;; CLASSIFICATION: <Unknown>  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/750,569  
;; FILING DATE: <Unknown>  
;; APPLICATION NUMBER: JP 7-120673  
;; FILING DATE: 21-APR-1995  
;; APPLICATION NUMBER: JP 6-311185  
;; FILING DATE: 21-NOV-1994  
;; APPLICATION NUMBER: JP 6-286917  
;; FILING DATE: 21-NOV-1994  
;; APPLICATION NUMBER: JP 6-290394  
;; FILING DATE: 31-OCT-1994  
;; APPLICATION NUMBER: JP 6-194223  
;; FILING DATE: 18-AUG-1994  
;; APPLICATION NUMBER: JP 6-133354  
;; FILING DATE: 16-JUN-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BENT, Stephen A.  
;; REGISTRATION NUMBER: 29,768  
;; REFERENCE/DOCKET NUMBER: 49441/110  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202)672-5300  
;; TELEFAX: (202)672-5399  
;; TELEX: 904136  
;;  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 559 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-298-924-6  
  
Query Match 76.5%; Score 26; DB 4; Length 559;  
Best Local Similarity 50.0%; Pred. No. 1.6e+02;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 EVMXVAEF 8  
Db 138 EIMPIAOF 145  
  
RESULT 3  
US-08-456-647B-18  
; Sequence 18, Application US/08456647B  
; Patent No. 5811516  
; GENERAL INFORMATION:  
; APPLICANT: Lemke Ph.D. et al., Greg E.  
; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/456,647B  
;; FILING DATE: 02-JUN-1995  
;; CLASSIFICATION: 530  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/237,401  
;; FILING DATE: 02-MAY-1994  
;; APPLICATION NUMBER: US 07/884,486  
;; FILING DATE: 15-MAY-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Wetherell Ph.D., John R.  
;; REGISTRATION NUMBER: 31,678  
;; REFERENCE/DOCKET NUMBER: 07251/007002  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (619) 678-5070  
;; TELEFAX: (619) 678-5099  
;; INFORMATION FOR SEQ ID NO: 18:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 54 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-456-647B-18  
  
Query Match 73.5%; Score 25; DB 2; Length 54;  
Best Local Similarity 50.0%; Pred. No. 24;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 EVMXVAEF 8  
Db 8 DVMKIADF 15  
  
RESULT 4  
US-08-237-401A-18  
; Sequence 18, Application US/08237401A  
; Patent No. 5837448  
; GENERAL INFORMATION:  
; APPLICANT: Lemke Ph.D. et al., Greg E.  
; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/237,401A  
; FILING DATE: 02-MAY-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/884,486  
; FILING DATE: 15-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile Ph.D., Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07251/007001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 678-5070  
; TELEFAX: (619) 678-5099  
; INFORMATION FOR SEQ ID NO: 18:



SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-237-401A-18

Query Match 73.5%; Score 25; DB 2; Length 54;  
Best Local Similarity 50.0%; Pred. No. 24;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 EVMXVAEF 8  
Db 8 DVMKIADF 15

## RESULT 5

US-09-291-023A-20  
Sequence 20, Application US/09291023A  
Patent No. 6309871

## GENERAL INFORMATION:

APPLICANT: Otttrup, Helle  
APPLICANT: Borchert, Torben  
APPLICANT: Nielsen, Bjarne  
APPLICANT: Nielsen, Vibeke  
APPLICANT: Hoeck, Lisbeth  
TITLE OF INVENTION: Polypeptides Having Alkaline Alpha-amylase Activity And Nucleic A  
TITLE OF INVENTION: Encoding Same  
FILE REFERENCE: 5821.010-US

CURRENT APPLICATION NUMBER: US/09/291.023A  
CURRENT FILING DATE: 1999-04-13  
PRIOR APPLICATION NUMBER: DK 1999 00438  
PRIOR FILING DATE: 1999-03-31

NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 20

LENGTH: 405  
TYPE: PRT

ORGANISM: Bacillus

US-09-291-023A-20

Query Match 73.5%; Score 25; DB 4; Length 405;  
Best Local Similarity 62.5%; Pred. No. 2e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EVMXVAEF 8  
Db 180 EMFAVAEF 187

## RESULT 6

US-08-446-803-2  
Sequence 2, Application US/08446803  
Patent No. 5824531

## GENERAL INFORMATION:

APPLICANT: Otttrup, Helle  
APPLICANT: Bisgaard-Frantzen, Henrik  
APPLICANT: Ostergaard, Peter Rabbek  
APPLICANT: Rasmussen, Michael Dolberg  
APPLICANT: Van Der Zee, Pia  
TITLE OF INVENTION: Alkaline Bacillus Amylase  
NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5824531o No. 5824531disk of No. 5824531th America  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446.803  
FILING DATE: 01-June-1995

## CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
REFERENCE/DOCKET NUMBER: 4157.204-US  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 867-0123

TELEFAX: (212) 878-9655

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 485 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-446-803-2

Query Match 73.5%; Score 25; DB 2; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EVMXVAEF 8

Db 260 EMFAVAEF 267

## RESULT 7

US-08-861-837-2  
Sequence 2, Application US/08861837  
Patent No. 5856164

## GENERAL INFORMATION:

APPLICANT: Otttrup, Helle  
APPLICANT: Bisgaard-Frantzen, Henrik  
APPLICANT: Ostergaard, Peter Rabbek  
APPLICANT: Rasmussen, Michael Dolberg  
APPLICANT: Van Der Zee, Pia

TITLE OF INVENTION: Alkaline Bacillus Amylase

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5856164o No. 5856164disk of No. 5856164th America  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: New York

COUNTRY: USA

ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/861.837

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/446.803

FILING DATE: 01-June-1995

ATTORNEY/AGENT INFORMATION:

NAME: Harrington, James J.

REGISTRATION NUMBER: 38,711

REFERENCE/DOCKET NUMBER: 4157.204-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 867-0123

TELEFAX: (212) 878-9655

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 485 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-861-837-2

Query Match 73.5%; Score 25; DB 2; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
I: |||||

Db 260 EMFAVAEF 267

## RESULT 8

US-08-600-656-2  
Sequence 2, Application US/08600656  
Patent No. 6093562  
GENERAL INFORMATION:  
APPLICANT: Bisgaard-Frantzen, Henrik  
APPLICANT: Svendsen, Allan  
APPLICANT: Borchert, Torben Vedel  
TITLE OF INVENTION: AMYLASE VARIANTS  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. 60935620 No. 6093562disk of No. 6093562th America, Inc.  
STREET: 405 Lexington Avenue, Suite 6400  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/600,656  
FILING DATE: 13-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4318,204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 485 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-600-656-2

Query Match 73.5%; Score 25; DB 3; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
I: |||||

Db 260 EMFAVAEF 267

## RESULT 9

US-09-170-670-2  
Sequence 2, Application US/09170670  
Patent No. 6187576  
GENERAL INFORMATION:  
APPLICANT: Svendsen, Allan  
APPLICANT: Borchert, Torben  
APPLICANT: Bisgaard-Frantzen Henrik  
TITLE OF INVENTION: Alpha-Amylase Mutants

FILE REFERENCE: 5276.200-US  
CURRENT APPLICATION NUMBER: US/09/170,670  
CURRENT FILING DATE: 1998-10-13  
EARLIER APPLICATION NUMBER: 1172/97  
EARLIER FILING DATE: 1997-10-13  
EARLIER APPLICATION NUMBER: 60/063,306  
EARLIER FILING DATE: 1997-10-28  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FastSEQ for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 485  
TYPE: PRT  
ORGANISM: Bacillus sp.  
US-09-170-670-2

Query Match 73.5%; Score 25; DB 4; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
I: |||||

Db 260 EMFAVAEF 267

## RESULT 10

US-09-170-670-8  
Sequence 8, Application US/09170670  
Patent No. 6187576  
GENERAL INFORMATION:  
APPLICANT: Svendsen, Allan  
APPLICANT: Borchert, Torben  
APPLICANT: Bisgaard-Frantzen Henrik  
TITLE OF INVENTION: Alpha-Amylase Mutants  
FILE REFERENCE: 5276.200-US  
CURRENT APPLICATION NUMBER: US/09/170,670  
CURRENT FILING DATE: 1998-10-13  
EARLIER APPLICATION NUMBER: 1172/97  
EARLIER FILING DATE: 1997-10-13  
EARLIER APPLICATION NUMBER: 60/063,306  
EARLIER FILING DATE: 1997-10-28  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FastSEQ for Windows Version 3.0  
SEQ ID NO 8  
LENGTH: 485  
TYPE: PRT  
ORGANISM: Bacillus sp.  
US-09-170-670-8

Query Match 73.5%; Score 25; DB 4; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
I: |||||

Db 260 EMFAVAEF 267

## RESULT 11

US-09-193-068-2  
Sequence 2, Application US/09193068  
Patent No. 6197565  
GENERAL INFORMATION:  
APPLICANT: Svendsen, Allan  
APPLICANT: Kjrulff, S ren  
APPLICANT: Bisgaard-Frantzen, Henrik  
APPLICANT: Andersen, Carsten  
TITLE OF INVENTION: -Amylase Variants  
FILE REFERENCE: 5709,000-US  
CURRENT APPLICATION NUMBER: US/09/193,068  
CURRENT FILING DATE: 1998-11-16  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: FastSEQ for Windows Version 3.0  
SEQ ID NO 2

;  
; LENGTH: 485  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
US-09-193-068-2

Query Match 73.5%; Score 25; DB 4; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVMXVAEF 8  
I: |||||  
Db 260 EMFAVAEF 267

RESULT 12  
US-09-193-068-8  
; Sequence 8, Application US/09193068  
; Patent No. 6197565  
; GENERAL INFORMATION:  
; APPLICANT: Svendsen, Allan  
; APPLICANT: Kjrulff, S ren  
; APPLICANT: Bisgaard-Frantzen, Henrik  
; APPLICANT: Andersen, Carsten  
; TITLE OF INVENTION: -Amylase Variants  
; FILE REFERENCE: 5709.000-US  
; CURRENT APPLICATION NUMBER: US/09/193,068  
; CURRENT FILING DATE: 1998-11-16  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 8  
; LENGTH: 485  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
US-09-193-068-8

Query Match 73.5%; Score 25; DB 4; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVMXVAEF 8  
I: |||||  
Db 260 EMFAVAEF 267

RESULT 13  
US-09-183-412-2  
; Sequence 2, Application US/09183412  
; Patent No. 6204232  
; GENERAL INFORMATION:  
; APPLICANT: Borchert, Torben V.  
; APPLICANT: Svendsen, Allan  
; APPLICANT: Andersen, Carsten  
; APPLICANT: Nielsen, Bjarne L.  
; APPLICANT: Nissen, Torben L.  
; APPLICANT: Kjaerulff, Soren  
; TITLE OF INVENTION: Alpha-Amulase Mutants  
; FILE REFERENCE: 5368.200-US  
; CURRENT APPLICATION NUMBER: US/09/183,412  
; CURRENT FILING DATE: 1998-10-30  
; EARLIER APPLICATION NUMBER: 60/064,662  
; EARLIER FILING DATE: 1997-11-06  
; EARLIER APPLICATION NUMBER: 60/093,234  
; EARLIER FILING DATE: 1998-07-17  
; EARLIER APPLICATION NUMBER: 1240/97  
; EARLIER FILING DATE: 1997-10-30  
; EARLIER APPLICATION NUMBER: PA 1998 00936  
; EARLIER FILING DATE: 1998-07-14  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 485  
; TYPE: PRT  
; ORGANISM: Bacillus sp.

US-09-183-412-2

Query Match 73.5%; Score 25; DB 4; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVMXVAEF 8  
I: |||||  
Db 260 EMFAVAEF 267

RESULT 14  
US-09-183-412-8  
; Sequence 8, Application US/09183412  
; Patent No. 6204232  
; GENERAL INFORMATION:  
; APPLICANT: Borchert, Torben V.  
; APPLICANT: Svendsen, Allan  
; APPLICANT: Andersen, Carsten  
; APPLICANT: Nielsen, Bjarne  
; APPLICANT: Nissen, Torben L.  
; APPLICANT: Kjaerulff, Soren  
; TITLE OF INVENTION: Alpha-Amulase Mutants  
; FILE REFERENCE: 5368.200-US  
; CURRENT APPLICATION NUMBER: US/09/183,412  
; CURRENT FILING DATE: 1998-10-30  
; EARLIER APPLICATION NUMBER: 60/064,662  
; EARLIER FILING DATE: 1997-11-06  
; EARLIER APPLICATION NUMBER: 60/093,234  
; EARLIER FILING DATE: 1998-07-17  
; EARLIER APPLICATION NUMBER: 1240/97  
; EARLIER FILING DATE: 1997-10-30  
; EARLIER APPLICATION NUMBER: PA 1998 00936  
; EARLIER FILING DATE: 1998-07-14  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 8  
; LENGTH: 485  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
US-09-183-412-8

Query Match 73.5%; Score 25; DB 4; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVMXVAEF 8  
I: |||||  
Db 260 EMFAVAEF 267

RESULT 15  
US-09-264-097-5  
; Sequence 5, Application US/09264097  
; Patent No. 6287826  
; GENERAL INFORMATION:  
; APPLICANT: No. 6287826man, Barrie Edmund  
; APPLICANT: Hendriksen, Hanne Vang  
; TITLE OF INVENTION: Enzymatic Preparation of Glucose Syrup  
; TITLE OF INVENTION: From Starch  
; FILE REFERENCE: 5278.200-US  
; CURRENT APPLICATION NUMBER: US/09/264,097  
; CURRENT FILING DATE: 1999-03-08  
; EARLIER APPLICATION NUMBER: PA 0321/98  
; EARLIER FILING DATE: 1998-03-09  
; EARLIER APPLICATION NUMBER: 60/079,209  
; EARLIER FILING DATE: 1998-03-24  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 485.  
; TYPE: PRT  
; ORGANISM: Bacillus

US-09-264-097-5

Query Match 73.5%; Score 25; DB 4; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
I: ||||  
Db 260 EMFAVAEF 267

Search completed: April 23, 2003, 12:46:30  
Job time : 14.9655 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:33:40 ; Search time 29.931 seconds  
(without alignments)  
48.188 Million cell updates/sec

Title: US-09-723-722a-78  
Perfect score: 29  
Sequence: 1 VMXVAEF 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	96.6	776	10 Q94IL7	Q94il7 populus tre
2	28	96.6	3414	12 Q9IG40	Q9ig40 langat viru
3	28	96.6	3414	12 Q9IG39	Q9ig39 langat viru
4	27	93.1	775	10 Q9FE38	Q9fe38 arabidopsis
5	27	93.1	802	10 Q9SUR2	Q9sur2 arabidopsis
6	27	93.1	814	10 Q8VX03	Q8vxq3 cymodocea n
7	26	89.7	262	17 Q8U0A9	Q8u0a9 pyrococcus
8	26	89.7	324	10 Q9ZPQ5	Q9zpq5 arabidopsis
9	26	89.7	821	2 Q9XBW4	Q9xbw4 porphyromon
10	25	86.2	284	17 Q9U3Z7	Q9u3z7 pyrococcus
11	25	86.2	1077	10 P92974	P92974 arabidopsis
12	24	82.8	149	17 Q8TLX8	Q8tlx8 methanosarc
13	24	82.8	165	5 Q8SRD1	Q8srd1 encephalico
14	24	82.8	195	16 Q97F65	Q97f65 clostridium
15	24	82.8	206	16 Q8XWK7	Q8xwk7 raistonia s
16	24	82.8	248	5 Q97008	Q97008 leishmania

17	24	82.8	249	16 Q9PKD5	Q9pkd5 chlamydia m
18	24	82.8	270	5 Q9NE74	Q9ne74 leishmania
19	24	82.8	317	5 Q9U375	Q9u375 caenorhabdi
20	24	82.8	367	10 Q8S0C0	Q8s0c0 oryza sativ
21	24	82.8	367	16 Q97KM0	Q97km0 clostridium
22	24	82.8	378	10 Q9FNG6	Q9fng6 arabidopsis
23	24	82.8	427	11 Q99LI3	Q99li3 mus musculu
24	24	82.8	439	16 Q8RG30	Q8rg30 fusobacteri
25	24	82.8	463	3 Q8WZN5	Q8wzn5 pleurotus o
26	24	82.8	484	10 Q9C7R2	Q9c7r2 arabidopsis
27	24	82.8	582	17 Q58843	Q58843 pyrococcus
28	24	82.8	584	17 Q9UZNO	Q9uzn0 pyrococcus
29	24	82.8	619	5 Q8SRA9	Q8sra9 encephalico
30	24	82.8	738	5 P91063	P91063 caenorhabdi
31	24	82.8	921	16 Q9K9V0	Q9k9v0 bacillus ha
32	24	82.8	1122	5 Q61460	Q61460 caenorhabdi
33	24	82.8	1158	5 Q965S4	Q965s4 caenorhabdi
34	23	79.3	56	13 Q91995	Q91995 xenopus lae
35	23	79.3	57	11 Q9QW56	Q9qw56 mus sp. pro
36	23	79.3	57	13 Q08556	Q08556 phasianidae
37	23	79.3	61	4 Q9P166	Q9p166 homo sapien
38	23	79.3	68	13 Q9YI21	Q9yi21 gallus gall
39	23	79.3	89	2 Q9ZGY2	Q9zgy2 yersinia pe
40	23	79.3	96	16 Q8U8Q1	Q8u8q1 agrobacteri
41	23	79.3	105	16 Q9L6S7	Q9l6s7 salmonella
42	23	79.3	112	16 Q8X7L7	Q8x7l7 escherichia
43	23	79.3	123	17 Q8TY74	Q8ty74 methanopyru
44	23	79.3	129	11 Q83000	Q83000 rattus norv
45	23	79.3	131	6 Q8WMP3	Q8wmp3 equus cabal

#### ALIGNMENTS

RESULT 1

Q94IL7 PRELIMINARY; PRT; 776 AA.  
AC Q94IL7;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Putative high-affinity potassium uptake transporter.  
GN KUP1.  
OS Populus tremula x Populus tremuloides.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids 1; Malpighiales; Salicaceae; Populus.  
OX NCBI\_TaxID=47664;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Langer K., Ache P., Fromm J., Hedrich R.;  
RT "PtKUP contributes to wood formation.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ299422; CAC39168.1;  
DR InterPro; IPR003855; K+\_transprtr.  
DR Pfam; PF02705; K\_trans; 1.  
DR TIGRFAMs; TIGR00794; kup; 1.  
SQ SEQUENCE 776 AA; 87303 MW; BEC03D57ED0869BC CRC64;

Query Match 96.6%; Score 28; DB 10; Length 776;  
Best Local Similarity 85.7%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7

DB 604 VMSVAEF 610

RESULT 2

Q9IG40 PRELIMINARY; PRT; 3414 AA.  
ID Q9IG40;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Polyprotein precursor.  
 OS Langat virus (strain TP21).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus.  
 OX NCBI\_TaxID=31638;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TP21;  
 RX MEDLINE=20192178; PubMed=10725214;  
 RA Campbell M.S., Pletnev A.G.;  
 RT "Infectious cDNA clones of Langat tick-borne flavivirus that differ  
 from their parent in peripheral neurovirulence.";  
 RL Virology 269:225-237(2000).  
 DR EMBL: AF253419; AAF75259.1; -  
 DR HSSP: P14336; 1SVB.  
 DR MEROPS: S07.001; -  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR001122; Flavi\_capsidC.  
 DR InterPro: IPR000336; Flavi\_glycoproteE.  
 DR InterPro: IPR001850; Flavi\_helicase.  
 DR InterPro: IPR000069; Flavi\_M.  
 DR InterPro: IPR001157; Flavi\_NS1.  
 DR InterPro: IPR000752; Flavi\_NS2A.  
 DR InterPro: IPR000487; Flavi\_NS2B.  
 DR InterPro: IPR001528; Flavi\_NS4B.  
 DR InterPro: IPR000208; Flavi\_NS5.  
 DR InterPro: IPR002335; Flavi\_propep.  
 DR InterPro: IPR002877; FtsJ.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR Pfam: PF01003; Flavi\_capsid; 1.  
 DR Pfam: PF00869; Flavi\_glycoprot; 1.  
 DR Pfam: PF02832; Flavi\_glycopC; 1.  
 DR Pfam: PF00949; Flavi\_helicase; 1.  
 DR Pfam: PF01004; Flavi\_M; 1.  
 DR Pfam: PF00948; Flavi\_NS1; 1.  
 DR Pfam: PF01005; Flavi\_NS2A; 1.  
 DR Pfam: PF01002; Flavi\_NS2B; 1.  
 DR Pfam: PF01350; Flavi\_NS4A; 1.  
 DR Pfam: PF01349; Flavi\_NS4B; 1.  
 DR Pfam: PF00972; Flavi\_NS5; 1.  
 DR Pfam: PF01570; Flavi\_propep; 1.  
 DR Pfam: PF01728; FtsJ; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR ProDom: PD001496; Flavi\_NS1; 1.  
 DR ProDom: PD001556; Flavi\_glycoprote; 1.  
 DR SMART: SM00490; HELIC; 1.  
 KW SIGNAL.  
 FT CHAIN 97 116 POTENTIAL.  
 FT CHAIN 1 96 CAPSID PROTEIN.  
 FT CHAIN 117 280 MEMBRANE-ASSOCIATED GLYCOPROTEIN  
 FT CHAIN 281 779 ENVELOPE MEMBRANE-ASSOCIATED  
 FT CHAIN 780 1128 GLYCOPROTEIN.  
 FT CHAIN 1129 1358 NONSTRUCTURAL PROTEIN NS1.  
 FT CHAIN 1359 1489 NONSTRUCTURAL PROTEIN NS2A.  
 FT CHAIN 1490 2110 NONSTRUCTURAL PROTEIN NS2B.  
 FT CHAIN 2111 2259 NONSTRUCTURAL PROTEIN NS3.  
 FT CHAIN 2260 2511 NONSTRUCTURAL PROTEIN NS4A.  
 FT CHAIN 2512 3414 NONSTRUCTURAL PROTEIN NS4B.  
 FT CHAIN 3414 AA; 378135 MW; 8DE86A46A0E8F7E3 CRC64;  
 SQ SEQUENCE

Query Match 96.6%; Score 28; DB 12; Length 3414;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 VMXVAEF 7

II IIII

Db 929 VMTVAEF 935

RESULT 3  
 O9IG39 PRELIMINARY; PRT; 3414 AA.  
 AC O9IG39; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Polyprotein precursor.  
 OS Langat virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus.  
 OX NCBI\_TaxID=11085;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATTENUATED STRAIN E5;  
 RX MEDLINE=20192178; PubMed=10725214;  
 RA Campbell M.S., Pletnev A.G.;  
 RT "Infectious cDNA clones of Langat tick-borne flavivirus that differ  
 from their parent in peripheral neurovirulence.";  
 RL Virology 269:225-237(2000).  
 DR EMBL: AF253420; AAF75260.1; -  
 DR HSSP: P14336; 1SVB.  
 DR MEROPS: S07.001; -  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR001122; Flavi\_capsidC.  
 DR InterPro: IPR000336; Flavi\_glycoprote.  
 DR InterPro: IPR001850; Flavi\_helicase.  
 DR InterPro: IPR000069; Flavi\_M.  
 DR InterPro: IPR001157; Flavi\_NS1.  
 DR InterPro: IPR000752; Flavi\_NS2A.  
 DR InterPro: IPR000487; Flavi\_NS2B.  
 DR InterPro: IPR000404; Flavi\_NS4A.  
 DR InterPro: IPR001528; Flavi\_NS4B.  
 DR InterPro: IPR000208; Flavi\_NS5.  
 DR InterPro: IPR002335; Flavi\_propep.  
 DR InterPro: IPR002877; FtsJ.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR Pfam: PF01003; Flavi\_capsid; 1.  
 DR Pfam: PF00869; Flavi\_glycoprot; 1.  
 DR Pfam: PF02832; Flavi\_glycopC; 1.  
 DR Pfam: PF00949; Flavi\_helicase; 1.  
 DR Pfam: PF01004; Flavi\_M; 1.  
 DR Pfam: PF00948; Flavi\_NS1; 1.  
 DR Pfam: PF01005; Flavi\_NS2A; 1.  
 DR Pfam: PF01002; Flavi\_NS2B; 1.  
 DR Pfam: PF01350; Flavi\_NS4A; 1.  
 DR Pfam: PF01349; Flavi\_NS4B; 1.  
 DR Pfam: PF00972; Flavi\_NS5; 1.  
 DR Pfam: PF01570; Flavi\_propep; 1.  
 DR Pfam: PF01728; FtsJ; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR ProDom: PD001496; Flavi\_NS1; 1.  
 DR ProDom: PD001556; Flavi\_glycoprote; 1.  
 DR SMART: SM00490; HELIC; 1.  
 KW SIGNAL.  
 FT CHAIN 97 116 POTENTIAL.  
 FT CHAIN 1 96 CAPSID PROTEIN.  
 FT CHAIN 117 280 MEMBRANE-ASSOCIATED GLYCOPROTEIN  
 FT CHAIN 281 779 ENVELOPE MEMBRANE-ASSOCIATED  
 FT CHAIN 780 1128 GLYCOPROTEIN.  
 FT CHAIN 1129 1358 NONSTRUCTURAL PROTEIN NS1.  
 FT CHAIN 1359 1489 NONSTRUCTURAL PROTEIN NS2A.  
 FT CHAIN 1490 2110 NONSTRUCTURAL PROTEIN NS2B.  
 FT CHAIN 2111 2259 NONSTRUCTURAL PROTEIN NS3.  
 FT CHAIN 2260 2511 NONSTRUCTURAL PROTEIN NS4A.  
 FT CHAIN 2512 3414 NONSTRUCTURAL PROTEIN NS4B.  
 FT CHAIN 3414 AA; 378014 MW; BF8F8ACEAB96D534 CRC64;  
 SQ SEQUENCE

Query Match 96.6%; Score 28; DB 12; Length 3414;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 DB 929 VMTVAEF 935

RESULT 4

Q9FE38 PRELIMINARY; PRT; 775 AA.

AC Q9FE38;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Tiny root hair 1 protein.  
 GN TRH1  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rigas S., Debrosses G., Haralampidis K., Vicente-Agullo F.,  
 RA Feldmann K.A., Grabov A., Dolan L., Hatzopoulos P.,  
 RT "A Potassium Transporter Required for Tip Growth in Arabidopsis.";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ296156; CAC16138.1;  
 DR EMBL: AJ296155; CAC16137.1;  
 DR InterPro: IPR003855; K+-transprtr.  
 DR Pfam: PF02705; K\_trans; 1.  
 DR TIGRFAMS: TIGR00794; kup; 1.  
 SQ SEQUENCE 775 AA; 86842 MW; B0C5068B48E8180 CRC64;

Query Match 93.1%; Score 27; DB 10; Length 775;  
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 DB 602 VMSIAEF 508

RESULT 5

Q9SUR2 PRELIMINARY; PRT; 802 AA.

AC Q9SUR2;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative potassium transport protein.  
 GN F9D16.110 OR AT4G23640.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Wedler H., Wedler E., Wambutt R., Hoheisel J., Mewes H.W.,  
 RA Mayer K.F.X., Schueller C.,  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Wedler H., Wedler E., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL035394; CAA23030.1;  
 DR EMBL: AL161559; CAB79319.1;  
 DR InterPro: IPR003855; K+-transprtr.  
 DR Pfam: PF02705; K\_trans; 1.  
 DR TIGRFAMS: TIGR00794; kup; 1.  
 SQ SEQUENCE 802 AA; 89817 MW; 617AF5F76B99BB60 CRC64;

Query Match 93.1%; Score 27; DB 10; Length 802;  
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 DB 629 VMSIAEF 635

RESULT 6

Q8VXQ3 PRELIMINARY; PRT; 814 AA.

AC Q8VXQ3;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative potassium transporter.  
 GN HAK2.  
 OS Cymodocea nodosa.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Cymodoceaceae; Cymodocea.  
 OX NCBI\_TaxID=55448;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Garcia-Abblas B., Benito B., Rodriguez-Navarro A.,  
 RT "Cloning and functional expression in bacteria of the potassium  
 RT transporters ChHAK1 and ChHAK2 from the seagrass Cymodocea nodosa.";  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ427293; CAD20319.1;  
 DR InterPro: IPR003855; K+-transprtr.  
 DR Pfam: PF02705; K\_trans; 1.  
 DR TIGRFAMS: TIGR00794; kup; 1.  
 SQ SEQUENCE 814 AA; 91609 MW; 6350453B857BEBBCB CRC64;

Query Match 93.1%; Score 27; DB 10; Length 814;  
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 DB 611 VMSIAEF 617

RESULT 7

Q8U0A9 PRELIMINARY; PRT; 262 AA.

AC Q8U0A9;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE 2-dehydro-3-deoxyphosphoheptomate aldolase (EC 4.1.2.15).  
 GN PF1690.  
 OS Pyrococcus furiosus.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococaceae; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxID=2261;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VCL / DSM 3638 / ATCC 43587 / JCM 8422;  
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.,  
 RT "The complete sequence of the Pyrococcus furiosus genome.";  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE010267; AAL81814.1;  
 KW Lyase; Complete proteome.  
 SQ SEQUENCE 262 AA; 29225 MW; ED84D28FD5DEADC2 CRC64;

```

Query Match      89.7%; Score 26; DB 17; Length 262;
Best Local Similarity 71.4%; Pred. No. 71;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7
   :| ||||
Db 39 IMKVAEF 45

RESULT 8
Q92PQ5 PRELIMINARY; PRT; 324 AA.
AC Q92PQ5;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE putative steroid sulfotransferase.
GN AT2G03770.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Roundsley S.D., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell L.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006836; AAD20079.1; -
DR HSSP; P50224; 1CJM.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
SQ SEQUENCE 324 AA; 37717 MW; B488C799AB13E1D3 CRC64;

Query Match      89.7%; Score 26; DB 10; Length 324;
Best Local Similarity 71.4%; Pred. No. 87;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7
   :| ||||
Db 233 VMRIAEF 239

RESULT 9
Q9XBW4 PRELIMINARY; PRT; 821 AA.
AC Q9XBW4;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Immunoreactive 92 kDa antigen Pg21.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; CFB group; Bacteroidetes; Bacteroidales; Porphyromonadaceae;
OC Porphyromonas.
OX NCBI_TaxID=837;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "Porphyromonas gingivalis polypeptides and nucleic acids.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF153768; AAD38980.1; -
DR InterPro; IPR000601; PKD_domain.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF00801; PKD; 3.
DR Pfam; PF00245; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00089; PKD; 3.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
DR PROSITE; PS00093; PKD; 3.
SQ SEQUENCE 821 AA; 91517 MW; 250843B2C9C833E2 CRC64;

Query Match      89.7%; Score 26; DB 2; Length 821;
Best Local Similarity 85.7%; Pred. No. 21e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7
   :| ||||
Db 565 VMPVAEF 571

RESULT 10
Q8U3Z7 PRELIMINARY; PRT; 284 AA.
AC Q8U3Z7;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Co-factor modifying protein.
GN PF0302.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=VCI / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE010154; AAL80426.1; -
KW Complete proteome.
SQ SEQUENCE 284 AA; 32512 MW; B19557B14451FFCD CRC64;

Query Match      86.2%; Score 25; DB 17; Length 284;
Best Local Similarity 57.1%; Pred. No. 1.4e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7
   :| ||||
Db 222 IMKIAEF 228

RESULT 11
P92974 PRELIMINARY; PRT; 1077 AA.
AC P92974;
DT 01-MAY-1997 (TRENBLrel. 03, Created)
DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Ubiquitin-activating enzyme El.
GN UBA1 OR UBA2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]

```



RP SEQUENCE FROM N.A.  
 RC STRAIN-ECOTYPE COLUMBIA;  
 RA Hatfield P.M., Carpenter, T.C., Vierstra R.D.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RX MEDLINE=98069011; PubMed=9405937;  
 RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.  
 RT Sequence features of the regions of 1,044,062 bp covered by thirteen  
 RT physically assigned P1 clones.";  
 RL DNA Res. 4:291-300(1997).  
 CC -!- FUNCTION: ACTIVATES UBIQUITIN BY FIRST ADENYLATING WITH ATP ITS  
 CC CARBOXY-TERMINAL GLYCINE RESIDUE AND THEREAFTER LINKING THIS  
 CC RESIDUE TO THE SIDE CHAIN OF A CYSTEINE RESIDUE IN EL, YIELDING AN  
 CC UBIQUITIN-EL THIOLESTER AND FREE AMP (BY SIMILARITY).  
 CC -!- PATHWAY: FIRST STEP IN UBIQUITIN CONJUGATION.  
 CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -!- SIMILARITY: TO EL IN OTHER SPECIES.  
 DR EMBL; U40566; AAB37569.1; -;  
 DR EMBL; AB006700; BAB08968.1; -;  
 DR InterPro; IPR000734; Lipase.  
 DR InterPro; IPR000205; NAD binding.  
 DR InterPro; IPR000594; Thif domain.  
 DR InterPro; IPR000127; Ubact\_repeat.  
 DR InterPro; IPR000011; Uqutin\_activ\_enz.  
 DR Pfam; PF00899; Thif; 2.  
 DR Pfam; PF02134; Ubact; 1.  
 DR PROSITE; PS00120; LIPASE\_SER; UNKNOWN\_1.  
 DR PROSITE; PS00536; UBIQUITIN\_ACTIVAT\_1; 1.  
 DR PROSITE; PS00865; UBIQUITIN\_ACTIVAT\_2; 1.  
 DR Ligase; Ubiquitin conjugation.  
 FT ACT\_SITE 653 653 BY SIMILARITY.  
 SQ SEQUENCE 1077 AA; 119622 MW; CE39A36AAA99A218 CRC64;

Query Match 86.2%; Score 25; DB 10; Length 1077;  
 Best Local Similarity 71.4%; Pred. No. 5e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 II : IIII  
 DB 54 VMSMAEF 60

RESULT 12  
 Q8TLY8 PRELIMINARY; PRT; 149 AA.  
 AC Q8TLY8;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical protein MA2884.  
 GN MA2884.  
 OS Methanosarcina acetivorans.  
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2214;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
 RX MEDLINE=21929760; PubMed=11932238;  
 RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P., Brown A.,  
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Ancoor D., Johnson R.,  
 RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,  
 RA Linton L., McKean P., McKernan K., Talamas J., Tirrell A., Ye W.,  
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,  
 RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,  
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
 RA Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,  
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,  
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,

RA Metcalf W.W., Birren B.;  
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic  
 RT and physiological diversity.";  
 RL Genome Res. 12:532-542(2002).  
 DR EMBL; AE010989; AAM06261.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 149 AA; 16402 MW; EF8576A410EC13F3 CRC64;

Query Match 82.8%; Score 24; DB 17; Length 149;  
 Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 II : IIII  
 DB 13 VLGVAEF 19

RESULT 13  
 Q8SRD1 PRELIMINARY; PRT; 165 AA.  
 AC Q8SRD1;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Coatomer zeta subunit.  
 GN ECU08\_0680.  
 OS Eucephalitozoon cuniculi.  
 OC Eukaryota; Microsporidia; Unikaryonidae; Eucephalitozoon.  
 OX NCBI\_TaxID=6035;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,  
 RA Prensier G., Barbe V., Peyretallade E., Brottier P., Wincker P.,  
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,  
 RA Weissbach J., Vivares C.P.;  
 RT "Genome sequence and gene compaction of the eukaryote parasite  
 RT Eucephalitozoon cuniculi.";  
 RL Nature 414:450-453(2001).  
 RL Nature 414:450-453(2001).  
 DR EMBL; AL590448; CAD26373.1; -;  
 SQ SEQUENCE 165 AA; 18742 MW; 7CDAB6C35BC11A6C CRC64;

Query Match 82.8%; Score 24; DB 5; Length 165;  
 Best Local Similarity 71.4%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 II : IIII  
 DB 107 VLLVAEF 113

RESULT 14  
 Q97F65 PRELIMINARY; PRT; 195 AA.  
 AC Q97F65;  
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Thymidine kinase.  
 GN CAC2887.  
 OS Clostridium acetobutylicum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostridiales; Clostridiaceae; Clostridium.  
 OX NCBI\_TaxID=1488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;

RX MEDLINE-21359325; PubMed-11466286;  
 RA Noelling J., Bréton G., Omeichenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
 RA Bennett G.N., Koonin E.V., Smith D.R.;  
 RT "Genome sequence and comparative analysis of the solvent-producing  
 RT bacterium *Clostridium acetobutylicum*."  
 RL J. Bacteriol. 183:4823-4838(2001).  
 DR EMBL: AE007786; AAK80830.1; -;  
 DR InterPro: IPR000345; Cytc\_heme\_bind.  
 DR InterPro: IPR001267; TK\_cell.  
 DR Pfam: PF00265; TK; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 DR PROSITE; PS00603; TK\_CELLULAR\_TYPE; UNKNOWN\_1.  
 KW Kinase; Complete proteome.  
 SQ SEQUENCE 195 AA; 22303 MW; 6EDB551986B203BA CRC64;

Query Match 82.8%; Score 24; DB 16; Length 195;  
 Best Local Similarity 57.1%; Pred. No. 1.9e+02;  
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7

Db 131 LMAIAEF 137

## RESULT 15

Q8XWK7 PRELIMINARY; PRT; 206 AA.  
 AC Q8XWK7;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Probable transmembrane protein.  
 GN RSC2467 OR RS01138.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GM11000;  
 RX MEDLINE-21681879; PubMed-11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*."  
 RL Nature 415:497-502(2002).  
 DR EMBL: AL646070; CAD16174.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 206 AA; 21783 MW; 8D49E45A150B8440 CRC64;

Query Match 82.8%; Score 24; DB 16; Length 206;  
 Best Local Similarity 57.1%; Pred. No. 2e+02;  
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7

Db 55 VLIAIAEF 61

Search completed: April 23, 2003, 12:38:39  
 Job time : 31.931 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:33:39 ; Search time 24.5517.Seconds  
(without alignments)  
43.419 Million cell updates/sec

Title: US-09-723-722A-81  
Perfect score: 34  
Sequence: 1 EVMXVAEF 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	94.1	8	21 AAB07872	A beta-secretase 1
2	27	79.4	7	21 AAB07871	A beta-secretase 1
3	27	79.4	347	22 AAG93168	C glutamic prote
4	27	79.4	802	23 AAB93074	Herbicidally activ
5	26	76.5	107	20 AAY35564	Chlamydia pneumoni
6	26	76.5	216	21 AAY95715	Cosmid CHRIM5 enco
7	26	76.5	218	22 AAB63744	Human prostate can
8	26	76.5	254	20 AAY35214	Amino acid sequenc
9	26	76.5	271	20 AAY15227	Human receptor pro
10	26	76.5	271	21 AAB28205	Novel human protei

11	26	76.5	271	22 AAG64357	Human signal recog
12	26	76.5	271	22 AAB92851	Human protein sequ
13	26	76.5	271	22 AAB88446	Human membrane or
14	26	76.5	271	23 ABB90282	Human polypeptide
15	26	76.5	276	21 AAB42611	Human ORFX ORF2375
16	26	76.5	282	21 AAB57116	Human prostate can
17	26	76.5	533	23 ABB47948	Listeria monocytog
18	26	76.5	559	17 AAR90619	Sulfolobus solfata
19	26	76.5	579	22 AAB96355	Putative p. abyssi
20	26	76.5	587	21 AAY74285	Neisseria meningit
21	26	76.5	587	21 AAY74287	Neisseria meningit
22	26	76.5	587	21 AAY74288	Neisseria meningit
23	26	76.5	781	23 AAU12049	Clostridium diffic
24	26	76.5	785	23 ABB97996	Human severing sti
25	26	76.5	821	20 AAY34479	Porphyromonas ging
26	26	76.5	869	20 AAY34354	Porphyromonas ging
27	26	76.5	1029	23 AAE17786	Human kinesin supe
28	26	76.5	1249	22 ABB71313	Drosophila melanog
29	25	73.5	9	23 ABB77871	A beta-amyloid pre
30	25	73.5	14	21 ABB07888	A peptide fragment
31	25	73.5	27	22 ABB28205	Human peptide #856
32	25	73.5	27	22 ABB33380	Peptide #886 encod
33	25	73.5	27	22 ABB18839	Protein #838 encod
34	25	73.5	27	22 AAM54165	Human bone marrow
35	25	73.5	27	22 AAM66559	Human brain expres
36	25	73.5	27	22 AAM14432	Peptide #866 encod
37	25	73.5	27	22 AAM26845	Peptide #882 encod
38	25	73.5	27	22 AAM02159	Peptide #841 encod
39	25	73.5	27	23 AAG36211	Human peptide enco
40	25	73.5	54	19 AAW79151	Receptor protein t
41	25	73.5	54	20 AAW81408	Propionibacterium
42	25	73.5	88	22 AAU56968	Novel human diagno
43	25	73.5	135	22 AAG26145	Novel human secret
44	25	73.5	144	22 AAG31774	Escherichia coli p
45	25	73.5	152	22 AAG98397	

## ALIGNMENTS

RESULT 1  
AAB07872  
ID AAB07872 standard; peptide; 8 AA.  
XX  
AC AAB07872;  
XX  
DT 14-NOV-2000 (first entry)  
XX  
DE A beta-secretase inhibitor peptide.  
XX  
KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KW inhibitor.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 4 /note= "hydroxyethylene"  
XX  
XX WO2000047618-A2.  
XX  
PD 17-AUG-2000.  
XX  
PF 10-FEB-2000; 2000WO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
PR 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

XX WPI; 2000-533011/48.  
 XX  
 PT Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease -  
 XX  
 XX  
 PS Disclosure; Page 12; 121pp; English.  
 XX  
 CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents an inhibitor of beta-secretase enzyme.  
 XX  
 SQ Sequence 8 AA;  
 Query Match 94.1%; Score 32; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 EVMXVAEF 8  
 Db 1 EVMXVAEF 8  
 RESULT 2  
 ID AAB07871 standard; peptide; 7 AA.  
 XX  
 AC AAB07871;  
 XX  
 DT 14-NOV-2000 (first entry)  
 XX  
 DE A beta-secretase inhibitor peptide.  
 XX  
 KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 KW inhibitor.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 3  
 FT /note= "hydroxyethylene"  
 XX  
 PN WO200047618-A2.  
 XX  
 PD 17-AUG-2000.  
 XX  
 PF 10-FEB-2000; 2000WO-US03819.  
 XX  
 XX 10-FEB-1999; 99US-0119571.  
 PR 15-JUN-1999; 99US-0139172.  
 PR  
 XX (ELAN-) ELAN PHARM INC.  
 XX  
 XX Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
 PI  
 XX WPI; 2000-533011/48.  
 DR  
 XX Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease -  
 XX  
 XX Disclosure; Page 12; 121pp; English.

CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents an inhibitor of beta-secretase enzyme.  
 XX  
 SQ Sequence 7 AA;  
 Query Match 79.4%; Score 27; DB 21; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 VMXVAEF 8  
 Db 1 VMXVAEF 7  
 RESULT 3  
 AAG93168  
 ID AAG93168 standard; Protein; 347 AA.  
 XX  
 AC AAG93168;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE C glutamicum protein fragment SEQ ID NO: 6922.  
 XX  
 DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 KW  
 KW Corynebacterium glutamicum.  
 XX  
 OS EPI108790-A2.  
 XX  
 PN 20-JUN-2001.  
 PD  
 XX 18-DEC-2000; 2000EP-0127688.  
 PF  
 XX 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 PR  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 PI  
 XX WPI; 2001-376931/40.  
 DR N-PSDB; AAH68387.  
 DR  
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX Claim 17; SEQ ID NO: 6922; 246pp + Sequence Listing; English.  
 PS  
 XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.

SQ Sequence 347 AA;

Query Match 79.4%; Score 27; DB 22; Length 347;  
 Best Local Similarity 62.5%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
 ::::|  
 Db 78 EILVAEF 85

#### RESULT 4

ABB93074  
 ID ABB93074 standard; Protein; 802 AA.

XX AC ABB93074;

XX DT 31-MAY-2002 (first entry)

XX DE Herbicidally active polypeptide SEQ ID NO 2285.

XX KW Herbicidal; plant; agriculture; herbicide.

XX OS Arabidopsis thaliana.

XX PN WO200210210-A2.

XX PD 07-FEB-2002.

XX PF 28-AUG-2001; 2001WO-EP09892.

XX PR 28-AUG-2001; 2001WO-EP09892.

XX PA (FARB ) BAYER AG.

XX PI Tietjen K, Weidner M;

XX WPI; 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms -

XX PS Claim 5; SEQ ID NO 2285; 261pp + Sequence Listing; English.

CC The invention relates to identifying target proteins  
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
 CC aligning and comparing nucleic acid or amino acid sequences from plant  
 CC with nucleic acid or amino acid sequences from non-plant organisms using  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are  
 CC useful as herbicides.

XX SQ Sequence 802 AA;

Query Match 79.4%; Score 27; DB 23; Length 802;  
 Best Local Similarity 71.4%; Pred. No. 5.3e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 VMXVAEF 8  
 ::::|  
 Db 629 VMSIAEF 635

#### RESULT 5

AAY35564  
 ID AAY35564 standard; Protein; 107 AA.

XX AAY35564;

XX DT 13-SEP-1999 (first entry)

XX DE Chlamydia pneumoniae protein not found in C. trachomatis.

XX KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope.

XX OS Chlamydia pneumoniae.

XX PN WO9927105-A2.

XX PD 03-JUN-1999.

XX PF 20-NOV-1998; 98WO-IB01890.

XX PR 04-NOV-1998; 98US-0107078.

XX PR 21-NOV-1997; 97FR-0014673.

XX PA (GEST ) GENSET.

XX PI Griffais R;

XX WPI; 1999-357842/30.

XX PT Genome sequence of Chlamydia pneumoniae

XX PS Page 1307; Disclosure; 1912pp; English.

XX AAY34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAY91990) of Chlamydia pneumoniae.  
 CC C. pneumoniae causes respiratory disease such as pneumonia and  
 CC bronchitis and is thought to be a contributing factor in heart  
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
 CC nucleotide sequences can also be used as immunogenic compositions,  
 CC especially where the vector directs the expression of a neutralising  
 CC epitope of C. pneumoniae.

XX SQ Sequence 107 AA;

Query Match 76.5%; Score 26; DB 20; Length 107;  
 Best Local Similarity 62.5%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
 ::::|  
 Db 91 EVMRIARF 98

#### RESULT 6

AAY95715

ID AAY95715 standard; Protein; 216 AA.

XX AC AAY95715;

XX DT 25-OCT-2000 (first entry)

XX DE Cosmid CHRIM5 encoded protein P21-7r.

XX KW Cosmid CHRIM5; nematocide; nematode; biological control agent;  
 KW transgenic plant; helminthiasis; P21-7r.

XX OS Xenorhabdus bovienii.

XX PN WO200042855-A1.

XX PD 27-JUL-2000.

XX 24-JAN-2000; 2000WO-GB00219.  
 XX  
 PF 22-JAN-1999; 99GB-0001499.  
 XX  
 PR (HORT-) HORTICULTURE RES INT.  
 XX  
 PA Morgan JAW, Jarrett P, Ellis D, Ousley MA;  
 PI  
 XX WPI: 2000-499157/44.  
 DR N-PSDB; AAA50029.  
 DR  
 XX Novel composition used to control parasitic nematodes, especially in  
 XX plants such as maize, cotton, soya, and rice, comprises a bacterium  
 XX PT which is a symbiont of an entomopathogenic nematode -  
 XX PT  
 XX Example 6; Page 44; 74pp; English.  
 PS  
 XX The present sequence is that of protein P21-7r encoded by an open  
 CC reading frame identified in cosmid CHRIM5 (see AAA50029). CHRIM5 was  
 CC obtained by ligating Xenorhabdus bovienii strain I73 (NCIMB 40986)  
 CC Sau3A-digested DNA fragments into the BamHI site of the Stratagene  
 CC cosmid vector SupercosI, packaging into Escherichia coli XL Blue I,  
 CC and screening for nematocidal activity against Caenorhabditis elegans.  
 CC Analysis of the DNA indicated a number of open reading frames for  
 CC which the corresponding protein sequences were determined (see  
 CC AA95685-Y95735). Nematodes can be controlled through the use of  
 CC bacteria associated symbiotically with an entomopathogenic nematode.  
 CC Such bacteria include Xenorhabdus and Photobacterium spp. such as X.  
 CC bovienii strain I73. The symbiont bacteria, an engineered  
 CC bacterium, or a nematocidal protein obtained from such bacteria,  
 CC can be used to control helminthiasis in a human or domesticated  
 CC animal or for the control of plant pathogen nematodes. Also  
 CC claimed are vectors for expressing nematocidal proteins in host  
 CC cells, and transgenic plants.  
 XX  
 XX SQ Sequence 216 AA;  
 Query Match 76.5%; Score 26; DB 21; Length 216;  
 Best Local Similarity 62.5%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 EVNXXVAEF 8  
 Db 108 EIMGVAF 115  
 RESULT 7  
 AAB63744  
 ID AAB63744 standard; Protein: 218 AA.  
 XX  
 AC AAB63744;  
 XX  
 XX 26-MAR-2001 (first entry)  
 DT  
 XX Human prostate cancer associated antigen protein sequence SEQ ID NO:1106.  
 DE  
 XX Human: breast cancer; gastric cancer; prostate cancer; diagnosis;  
 XX KW cancer associated antigen; cytostatic; cancer vaccine.  
 XX KW  
 XX Homo sapiens.  
 OS  
 XX WO200073801-A2.  
 PN  
 XX 07-DEC-2000.  
 PD  
 XX 26-MAY-2000; 2000WO-US14749.  
 PF  
 XX 28-MAY-1999; 99US-0136526.  
 PR  
 XX 10-SEP-1999; 99US-0153454.  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 XX

PI Obata Y;  
 XX WPI: 2001-025274/03.  
 XX  
 PT Nucleic acids encoding breast, gastric and prostate cancer associated  
 PT antigen precursors, useful for diagnosing and treating a condition  
 PT characterized by expression of an abnormal amount of a protein, e.g.  
 PT cancer -  
 XX  
 PS Example 1; Page 696-697; 799pp; English.  
 PS  
 XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014  
 CC represent nucleotide sequences encoding human breast, gastric and  
 CC prostate cancer associated antigen precursors (CAAP) respectively.  
 CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970  
 CC represent human breast, gastric and prostate CAAP protein sequence  
 CC respectively. CAAPs have cytostatic activity and can be used in the  
 CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic  
 CC acids or anti-CAAP antibodies are useful for diagnosing and treating a  
 CC condition characterised by expression of an abnormal amount of a protein,  
 CC e.g. cancer.  
 XX  
 XX SQ Sequence 218 AA;  
 Query Match 76.5%; Score 26; DB 22; Length 218;  
 Best Local Similarity 75.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 EVNXXVAEF 8  
 Db 167 EVKNVAEF 174  
 RESULT 8  
 AAY35214  
 ID AAY35214 standard; Protein: 254 AA.  
 XX  
 AC AAY35214;  
 XX  
 XX 13-SEP-1999 (first entry)  
 DT  
 XX Amino acid sequence of a Chlamydia pneumoniae protein.  
 DE  
 XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 XX KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 XX KW vaccine; neutralising epitope.  
 XX  
 OS Chlamydia pneumoniae.  
 OS  
 XX WO9927105-A2.  
 PN  
 XX 03-JUN-1999.  
 PD  
 XX 20-NOV-1998; 98WO-IB01890.  
 PF  
 XX 04-NOV-1998; 98US-0107078.  
 PR  
 XX 21-NOV-1997; 97FR-0014673.  
 PA (GEST ) GENSET.  
 XX  
 XX Griffais R;  
 PI  
 XX WPI: 1999-357842/30.  
 DR  
 XX Genome sequence of Chlamydia pneumoniae  
 PT  
 XX Page 1062-1063; Disclosure: 1912pp; English.  
 PS  
 XX AAY34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.  
 CC C. pneumoniae causes respiratory disease such as pneumonia and  
 CC bronchitis and is thought to be a contributing factor in heart  
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema

CC nodusum or pharyngitis. The polypeptides encoded by the open reading  
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.  
XX  
SQ Sequence 254 AA;

Query Match 76.5%; Score 26; DB 20; Length 254;  
Best Local Similarity 50.0%; Pred. No. 2.7e+02;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVMXVAEF 8  
I:::||||  
DB 204 ELLAIAEF 211

RESULT 9  
AAY15227  
ID AAY15227 standard; protein; 271 AA.

XX AAY15227;

AC AAY15227;

DT 26-OCT-1999 (first entry)

DE Human receptor protein (HURP) 6 amino acid sequence.

XX receptor; cancer; autoimmune disorder; inflammation;  
KW antagonist; cell surface protein; cell signalling;  
KW antibody; human receptor protein; HURP; reproductive disorder;  
KW developmental disorder; gastrointestinal disorder.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 63 /note= "Potential CAMP-/cGMP-dependent protein-  
FT kinase phosphorylation site"

FT Misc-difference 95 /note= "Potential casein kinase II-  
FT phosphorylation site"

FT Misc-difference 114 /note= "Potential casein kinase II-  
FT phosphorylation site"

FT Misc-difference 213 /note= "Potential casein kinase II-  
FT phosphorylation site"

FT Misc-difference 6 /note= "Potential protein kinase C-  
FT phosphorylation site"

FT Misc-difference 25 /note= "Potential protein kinase C-  
FT phosphorylation site"

FT Misc-difference 59 /note= "Potential protein kinase C-  
FT phosphorylation site"

FT Misc-difference 63 /note= "Potential protein kinase C-  
FT phosphorylation site"

FT Misc-difference 75 /note= "Potential protein kinase C-  
FT phosphorylation site"

FT Misc-difference 123 /note= "Potential protein kinase C-  
FT phosphorylation site"

FT Misc-difference 135 /note= "Potential protein kinase C-  
FT phosphorylation site"

FT Misc-difference 189 /note= "Potential protein kinase C-  
FT phosphorylation site"

FT Misc-difference 203

FT /note= "Potential protein kinase C-  
FT phosphorylation site"  
FT Misc-difference 247 /note= "Potential protein kinase C-  
FT phosphorylation site"  
FT Misc-difference 160 /note= "Potential tyrosine kinase-  
FT phosphorylation site"  
FT Binding-site 71..78 /note= "Potential ATP/GTP-binding site-  
FT motif A (P-loop)"

XX WO9941375-A2.

PD 19-AUG-1999.

XX 05-FEB-1999; 99WO-US02572.

XX 12-FEB-1998; 98US-0022939.

XX (INCY-) INCYTE PHARM INC.

XX Au-Young J, Bandman O, Baughn M, Corley NC, Guegler KJ;  
PI Hillman JL, Lal P, Shah P, Tang YT, Yue H;

DR WPI: 1999-494536/41.  
DR N-PSDB; AAZ06371.

XX New human receptor proteins, used e.g. to treat, prevent and  
FT diagnose gastrointestinal and developmental disorders - and related  
FT nucleic acids, vectors, transformed cells, antibodies, agonists and  
FT antagonists

PS Claim 1; Page 83; 94pp; English.

XX The Human receptor protein 6 (HURP-6) and mouse signal recognition  
CC particle beta subunit share 90% identity.  
CC HURP-6 is expressed in cancerous, inflamed, reproductive and gastro-  
CC intestinal tissue. HURP-6 therefore appears to have a role in  
CC cancer, autoimmune/inflammatory disorders, reproductive disorders,  
CC and gastrointestinal disorders.

CC This gives rise to the possibility of using an antagonist or an antibody  
CC of HURP to treat or prevent cancer or autoimmune/inflammatory disorders.

SQ Sequence 271 AA;

Query Match 76.5%; Score 26; DB 20; Length 271;  
Best Local Similarity 75.0%; Pred. No. 2.9e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVMXVAEF 8

II:::||||

DB 151 EVKDVAEF 158

RESULT 10

AAB28205  
ID AAB28205 standard; Protein; 271 AA.

XX AAB28205;

XX 30-JAN-2001 (first entry)

XX Novel human protein #3.

DE Cytostatic; vaccine; human; breast tumour; antigen; breast cancer.

XX Homo sapiens.

XX WO200052165-A2.

XX 08-SEP-2000.

XX

PF 29-FEB-2000; 2000WO-US05431.

XX 04-MAR-1999; 99US-0262505.

PR 19-MAR-1999; 99US-0272886.

PR 17-SEP-1999; 99US-0396313.

XX (CORI-) CORIXA CORP.

PA

XX

XX

PI Lodes MJ;

XX

XX

DR WPI; 2000-572184/53.

DR N-PSDB; AAC69684.

XX Breast tumor antigen polypeptides and polynucleotides, useful for

PT manufacturing vaccines and compositions for treating, diagnosing, and

PT monitoring breast cancer.

XX Example; Fig 2; 140pp; English.

PS

XX

XX

CC The present invention relates to immunogenic portions of new human

CC breast tumor antigens (AAB28183-B28214) and their coding sequences

CC (AAC69645-C69804). The breast tumor antigen polypeptides of the present

CC invention and their coding sequences are useful for inhibiting the

CC development of breast cancer in a patient. The breast tumor antigen

CC polypeptides and polynucleotides may be used in vaccines and

CC pharmaceutical compositions for treating breast cancer, and for

CC diagnosing and monitoring the cancer. The present sequence is a

CC immunogenic portion for one such human breast cancer tumour antigen.

XX

SQ Sequence 271 AA:

Query Match 76.5%; Score 26; DB 21; Length 271;

Best Local Similarity 75.0%; Pred. No. 2.9e+02;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVNXXVAEF 8

II IIII

Db 151 EVKDVAEF 158

RESULT 11

AAG64357

ID AAG64357 standard; Protein; 271 AA.

XX AAG64357;

AC

XX 01-OCT-2001 (first entry)

DT Human signal recognition particle receptorbeta.

DE Human; signal recognition particle receptor beta; SRPRbeta.

XX Homo sapiens.

XX CNI279290-A.

PN

XX

PD 10-JAN-2001.

XX

XX

PF 23-JUN-1999; 99CN-0108547.

XX

XX

PR 23-JUN-1999; 99CN-0108547.

XX

XX (UYFU-) UNIV FUDAN.

PA

XX

XX

PI Yu L, Fu Q, Zhao Y;

XX

XX

DR WPI; 2001-266742/28.

DR N-PSDB; AAH73878.

XX New human signal recognition particle receptor beta nucleic acid for

PT preparing the protein encoded by it.

PT

PS Claim 2; Pages 15-16 (Disclosure); 20pp; Chinese.

XX

CC The present sequence is the protein sequence for human signal recognition

CC particle receptorbeta (SRPRbeta). The present protein is the homolog of

CC mouse SRPR beta. Application of human SRPRbeta coding sequence and

CC protein, and their preparing process are also disclosed.

XX

SQ Sequence 271 AA:

Query Match 76.5%; Score 26; DB 22; Length 271;

Best Local Similarity 75.0%; Pred. No. 2.9e+02;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVNXXVAEF 8

II IIII

Db 151 EVKDVAEF 158

RESULT 12

AAB92851

ID AAB92851 standard; Protein; 271 AA.

XX

AC AAB92851;

XX

DT 26-JUN-2001 (first entry)

XX

XX Human protein sequence SEQ ID NO:11414.

DE

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

KW

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

XX

PF 28-JUL-2000; 2000EP-0116126.

XX

XX

PR 29-JUL-1999; 99JP-0248036.

XX

PR 27-AUG-1999; 99JP-0300253.

XX

PR 11-JAN-2000; 2000JP-0118776.

XX

PR 02-MAY-2000; 2000JP-0183767.

XX

PR 09-JUN-2000; 2000JP-0241899.

XX

XX (HELI-) HELIX RES INST.

PA

XX

XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

DR WPI; 2001-318749/34.

XX

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs.

XX

XX

PS Claim 8; SEQ ID 11414; 2537pp + CD ROM; English.

XX

XX

CC The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by



CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX SQ Sequence 271 AA;

Query Match 76.5%; Score 26; DB 22; Length 271;  
 Best Local Similarity 75.0%; Pred. No. 2.9e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
 || ||||  
 Db 151 EVKDVAEF 158

#### RESULT 13

AAB88446  
 ID AAB88446 standard; Protein; 271 AA.

XX AC AAB88446;

XX DT 23-MAY-2001 (first entry)

XX DE Human membrane or secretory protein clone PSEC0230.

XX KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
 KW Rheumatoid arthritis; diabetes.

XX OS Homo sapiens.

XX PN EP1067182-A2.

XX PD 10-JAN-2001.

XX PF 07-JUL-2000; 2000EP-0114090.

XX PR 08-JUL-1999; 95JP-0194179.

XX PR 11-JAN-2000; 2000JP-0118775.

XX PR 02-MAY-2000; 2000JP-0183766.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
 XX WPI: 2001-093989/11.

XX DR N-PSDB; AAF93873.

XX PT Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 XX gene therapy or as candidate target molecules in drug development -

XX PS Claim 1; SEQ ID 260; 609pp + CD ROM; English.

XX CC This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by  
 CC AAB88317 - AAB88419. Included in the invention are primers  
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the  
 CC cDNA sequences of the invention. The invention also includes methods for  
 CC the production of antibodies directed against the proteins, and cDNA  
 CC sequences, which can be used in vaccines. The polynucleotide sequences  
 CC can be used in gene therapy. The polynucleotide sequences and the  
 CC proteins they encode may be used in the prevention, treatment and  
 CC diagnosis of diseases associated with inappropriate secretory  
 CC protein/membrane protein expression. The nucleic acids and complementary  
 CC sequences may also be used as DNA probes in diagnostic assays  
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the  
 CC presence of similar nucleic acid sequences in samples. They may also be  
 CC used to study the expression and function of secretory proteins/membrane  
 CC polypeptides and their role in metabolism. The polypeptides may be used  
 CC as antigens in the production of antibodies against them and in assays to  
 CC identify modulators (agonists and antagonists) of expression and

CC activity. The antibodies and antagonists may also be used as therapeutic  
 CC agents to down regulate expression and activity. The antibodies may also  
 CC be used as diagnostic agents for detecting the presence of the  
 CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay  
 CC (ELISA). Examples of diseases which may be treated include rheumatoid  
 CC arthritis and diabetes.

XX SQ Sequence 271 AA;

Query Match 76.5%; Score 26; DB 22; Length 271;  
 Best Local Similarity 75.0%; Pred. No. 2.9e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
 || ||||  
 Db 151 EVKDVAEF 158

#### RESULT 14

ABB90282  
 ID ABB90282 standard; Protein; 271 AA.

XX AC ABB90282;

XX DT 24-MAY-2002 (first entry)

XX DE Human polypeptide SEQ ID NO 2658.

XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.

XX OS Homo sapiens.

XX PN WO200190304-A2.

XX PD 29-NOV-2001.

XX PF 18-MAY-2001; 2001WO-US16450.

XX PR 19-MAY-2000; 2000US-205515P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Birse CE, Rosen CA;

XX DR WPI; 2002-122018/16.

XX DR N-PSDB; ABL90691.

XX PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 XX prevention of neural, immune system, muscular, cardiovascular, renal and proliferative  
 XX disorders -

XX PS Claim 11; SEQ ID NO 2658; 2081pp + Sequence Listing; English.

XX CC The invention relates to novel genes (ABL89449-ABL90853) and proteins  
 CC (ABB9040-ABB90444) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.

CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 271 AA;

Query Match 76.5%; Score 26; DB 23; Length 271;  
 Best Local Similarity 75.0%; Pred. No. 2.9e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVNXXVAEF 8  
 || ||||  
 Db 151 EVKDVAEF 158

# RESULT 15

AAB42611  
 ID AAB42611 standard; Protein; 276 AA.

XX AC AAB42611;

XX 08-FEB-2001 (first entry)

DE Human ORFX ORF2375 polypeptide sequence SEQ ID NO:4750.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnery; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antihypertensive;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shinkets RA, Leach M;

XX WPI; 2000-602362/57.

XX N-PSDB; AAC76820.

XX Novel nucleic acids and peptides derived from open reading frame X,

XX useful for treating e.g. cancers, proliferative disorders,

XX neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 3927; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

XX which represent the human ORFX open reading frames 1 to 3161. The ORFX

XX sequences have activities such as: cytostatic; hepatotropic; vulnery;

XX antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;

XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;

CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX SQ Sequence 276 AA;

Query Match 76.5%; Score 26; DB 21; Length 276;  
 Best Local Similarity 75.0%; Pred. No. 3e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVNXXVAEF 8

|| ||||

Db 156 EVKDVAEF 163

Search completed: April 23, 2003, 12:35:19

Job time : 25.5517 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:33:40 ; Search time 8.44828 seconds  
(without alignments)  
79.654 Million cell updates/sec

Title: US-09-723-722a-78  
Perfect score: 29  
Sequence: 1 VMXVAEF 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*

1: piri:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	96.6	2638	1 A42545	genome polyprotein
2	27	93.1	802	2 T05596	probable potassium
3	26	89.7	223	2 E64205	deoxyribose-phosph
4	26	89.7	224	2 S02216	deoxyribose-phosph
5	26	89.7	324	2 B84452	probable sterold s
6	25	86.2	201	2 F69988	hypothetical prote
7	25	86.2	632	2 T38126	probable electron
8	24	82.8	145	1 B41715	ribosomal protein
9	24	82.8	195	2 C97255	thymidine kinase (
10	24	82.8	236	2 B70728	hypothetical prote
11	24	82.8	248	2 T18315	hypothetical prote
12	24	82.8	249	2 G81693	serine/threonine p
13	24	82.8	335	2 F95103	6-phosphofructokin
14	24	82.8	335	2 D97971	6-phosphofructokin
15	24	82.8	367	2 F97010	fusion, chorismate
16	24	82.8	484	2 E86416	unknown protein, 3
17	24	82.8	565	2 B72660	probable type II D
18	24	82.8	582	2 E71052	hypothetical prote
19	24	82.8	584	2 F75090	archaeosine trna-r
20	24	82.8	734	2 T25551	hypothetical prote
21	24	82.8	919	2 T29581	hypothetical prote
22	24	82.8	921	2 H69643	isoleucine-tRNA li
23	24	82.8	921	2 A83968	isoleucyl-tRNA syn
24	24	82.8	1122	2 T42400	Eph receptor tyros
25	23	79.3	89	2 T15018	hypothetical prote
26	23	79.3	96	2 A83053	hypothetical prote
27	23	79.3	112	2 E90987	hypothetical prote
28	23	79.3	112	2 H85832	hypothetical prote
29	23	79.3	143	2 H69515	riboflavin synthas

30 23 79.3 157 2 G00016  
31 23 79.3 162 2 G86842  
32 23 79.3 169 2 AB2739  
33 23 79.3 172 2 S27019  
34 23 79.3 172 2 S27020  
35 23 79.3 172 2 S27021  
36 23 79.3 172 2 S27022  
37 23 79.3 187 2 S16314  
38 23 79.3 191 2 A82392  
39 23 79.3 209 2 H97519  
40 23 79.3 248 2 D81436  
41 23 79.3 254 1 BVB553  
42 23 79.3 275 2 S28749  
43 23 79.3 291 2 G69479  
44 23 79.3 370 2 F36819  
45 23 79.3 373 2 AC0253

## ALIGNMENTS

### RESULT 1

A42545

genome polyprotein - Langkat virus (strain tp21) (fragment)

N;Contains: nonstructural protein NS1; nonstructural protein NS2a; nonstructural protein NS5

C;Species: Langkat virus

C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 19-Jan-2001

C;Accession: A42545; A61409; G61409

R;Iacono-Conners, L.C.; Schmaljohn, C.S.

Virology 188, 875-880, 1992

A;Title: Cloning and sequence analysis of the genes encoding the nonstructural proteins

A;Reference number: A42545; MUID:92263794; PMID:1316684

A;Accession: A42545

A;Molecule type: genomic RNA

A;Residues: 1-2638 <IAC>

A;Cross-references: GB:S35365; NID:g249315; PIDN:AAB22165.1; PID:g249316

R;Guirakhoov, F.; Heinz, F.X.; Mandl, C.W.; Holzmann, H.; Kunz, C.; Gresikova, M.

J. Gen. Virol. 72, 333-338, 1991

A;Title: The relationship between the flaviviruses Stalica and Langkat as revealed by

A;Reference number: A61409; MUID:91132129; PMID:1847173

A;Accession: A61409

A;Status: not compared with conceptual translation

A;Molecule type: genomic RNA

A;Residues: 319-337 <GUI>

A;Accession: G61409

A;Status: not compared with conceptual translation

A;Molecule type: genomic RNA

A;Residues: 877-994 <GU2>

C;Superfamily: yellow fever virus genome polyprotein

C;Keywords: glycoprotein; nonstructural protein; nucleotide binding; P-loop; polypro-

F;1-352/Product: nonstructural protein NS1 #status predicted <NS1>

F;353-582/Product: nonstructural protein NS2a #status predicted <N2a>

F;583-713/Product: nonstructural protein NS2b #status predicted <N2b>

F;714-1334/Product: nonstructural protein NS3 #status predicted <NS3>

F;912-919/Region: nucleotide-binding motif A (P-loop)

F;1335-1463/Product: nonstructural protein NS4a #status predicted <N4a>

F;1484-1735/Product: nonstructural protein NS4b #status predicted <N4b>

F;1736-2638/Product: nonstructural protein NS5 #status predicted <NS5>

F;85,207,223,873,1212,1671,1950/Binding site: carbohydrate (Asn) (covalent) #status i

Query Match 96.6%; Score 28; DB 1; Length 2638;

Best Local Similarity 85.7%; Pred. No. 78;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7

Db 153 VMTVAEF 159.

### RESULT 2

T05596

probable potassium transport protein F9D16.110 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 18-Aug-2000  
 C:Accession: T05596  
 R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.  
 submitted to the Protein Sequence Database, February 1999  
 A:Reference number: Z15419  
 A:Accession: T05596  
 A:Molecule type: DNA  
 A:Residues: 1-802 <BRV>  
 A:Cross-references: EMBL:AL035394  
 A:Experimental source: cultivar Columbia; BAC clone F9D16  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 1/3; 16/3; 102/1; 185/1; 209/1; 296/1; 313/3; 352/1; 437/1  
 A>Note: F9D16.110  
 C:Superfamily: barley probable potassium transport protein HAK1  
 C:Keywords: ion transport

Query Match 93.1%; Score 27; DB 2; Length 802;  
 Best Local Similarity 71.4%; Pred. No. 43;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 II : III  
 Db 629 VMSTAEF 635

RESULT 3  
 E64205  
 deoxyribose-phosphate aldolase (EC 4.1.2.4) - Mycoplasma genitalium  
 C:Species: Mycoplasma genitalium  
 C:Date: 17-Nov-1995 #sequence\_revision 17-Nov-1995 #text\_change 21-Jul-2000  
 C:Accession: E64205  
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;  
 M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.  
 , C.A.; Venter, J.C.  
 Science 270, 397-403, 1995  
 A:Title: The minimal gene complement of Mycoplasma genitalium.  
 A:Reference number: A64200; MUID:96026346; PMID:7569993  
 C:Accession: E64205  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-223 <TRIGR>  
 A:Cross-references: GB:U39684; GB:U43967; NID:g3844650; PIDN:AAC71266.1; PID:g1045723; T  
 A:Experimental source: strain G-37  
 C:Genetics:  
 A:Genetic code: SGC3  
 C:Superfamily: deoxyribose-phosphate aldolase  
 C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 89.7%; Score 26; DB 2; Length 223;  
 Best Local Similarity 71.4%; Pred. No. 21;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 II : III  
 Db 93 VMSTAEF 99

RESULT 4  
 S02216  
 deoxyribose-phosphate aldolase (EC 4.1.2.4) - Mycoplasma pneumoniae  
 C:Species: Mycoplasma pneumoniae  
 A:Variety: ATCC 29342  
 C:Date: 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 07-Dec-1999  
 C:Accession: S02216; S73417  
 R:Loechel, S.; Inamine, J.M.; Hu, P.C.  
 Nucleic Acids Res. 17, 801, 1989  
 A:Title: Nucleotide sequence of the deoC gene of Mycoplasma pneumoniae.  
 A:Reference number: S02216; MUID:89128453; PMID:2492658  
 A:Accession: S02216  
 A:Molecule type: DNA  
 A:Residues: 1-224 <LOE>

A:Cross-references: EMBL:X13544; NID:g44480; PIDN:CAA31897.1; PID:g44481  
 R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.  
 Nucleic Acids Res. 24, 4420-4449, 1996  
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma  
 A:Reference number: S73327; MUID:97105885; PMID:8948633  
 A:Accession: S73417  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-224 <HIM>  
 A:Cross-references: EMBL:AE000011; GB:U00089; NID:g1673740; PIDN:AAB95739.1; PID  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November  
 C:Genetics:  
 A:Genetic code: SGC3  
 C:Superfamily: deoxyribose-phosphate aldolase  
 C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 89.7%; Score 26; DB 2; Length 224;  
 Best Local Similarity 71.4%; Pred. No. 22;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 II : III  
 Db 93 VMSTAEF 99

RESULT 5  
 B84452  
 probable steroid sulfotransferase [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
 C:Accession: B84452  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; T  
 euss, D.; Nierman, W.C.; White, O.; Eelsen, J.A.; Salzberg, S.L.; Fraser, C.M.;  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 C:Accession: B84452  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-324 <STO>  
 A:Cross-references: GB:AE002093; NID:g4406768; PIDN:AAD20079.1; GSPDB:GN00139  
 C:Genetics:  
 A:Map position: 2  
 C:Superfamily: alcohol sulfotransferase

Query Match 89.7%; Score 26; DB 2; Length 324;  
 Best Local Similarity 71.4%; Pred. No. 31;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 II : III  
 Db 233 VMSTAEF 239

RESULT 6  
 F69988  
 hypothetical protein ytbQ - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
 C:Accession: F69988  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.  
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter  
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Fe  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Puma, S.; Galizzi, A  
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; H  
 koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.;  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S  
 y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; P  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: F69988  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-201 <KUN>  
A:Cross-references: GB:299119; GB:AL009126; NID:g2635411; PIDN:CAB14996.1; PID:ell185891.  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ytbQ

Query Match 86.2%; Score 25; DB 2; Length 201;  
Best Local Similarity 57.1%; Pred. No. 36;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VMXVAEF 7  
| : |||  
Db 23 IMDIAEF 29

## RESULT 7

T38126  
probable electron transfer flavoprotein precursor - fission yeast (*Schizosaccharomyces* sp.)  
C:Species: *Schizosaccharomyces pombe*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T38126  
R:Badcock, K.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1997  
A:Reference number: 221772  
A:Accession: T38126  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-632 <BAD>  
A:Cross-references: EMBL:Z95334; PIDN:CA808598.1; GSPDB:GN00066; SPDB:SPAC2068.04c  
A:Experimental source: strain 972h-; cosmid c2068  
C:Genetics:  
A:Gene: SPDB:SPAC2068.04c  
A:Map position: 1

Query Match 86.2%; Score 25; DB 2; Length 632;  
Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VMXVAEF 7  
| : |||  
Db 193 VMSLAEF 199

## RESULT 8

B41715  
ribosomal protein L13 [similarity] - *Halobacterium marismortui*  
C:Species: *Halobacterium marismortui*  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
C:Accession: B41715  
R:Kroemer, W.J.; Arndt, E.  
J. Biol. Chem. 266, 24573-24579, 1991  
A:Title: *Halobacterium* S9 operon. Three ribosomal protein genes are cotranscribed with a  
bacterial *marismortui*.  
A:Reference number: A41715; MUID:92105119; PMID:1840597  
A:Accession: B41715  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-145 <KRO>  
A:Cross-references: GB:M76567; NID:g148775; PIDN:AAA73097.1; PID:g148777  
C:Superfamily: *Escherichia coli* ribosomal protein L13

Query Match 82.8%; Score 24; DB 1; Length 145;  
Best Local Similarity 83.3%; Pred. No. 48;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 MXVAEF 7  
| : |||  
Db 1 MSVAEF 6

## RESULT 9

C97255  
thymidine kinase (EC 2.7.1.21) [similarity] - *Clostridium acetobutylicum*  
C:Species: *Clostridium acetobutylicum*  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 31-Dec-2001  
C:Accession: C97255  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L  
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: C97255  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-195 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK80830.1; PID:g15025935; GSPDB:GN00168  
A:Experimental source: *Clostridium acetobutylicum* ATCC824  
C:Genetics:  
A:Gene: CAC2887  
C:Superfamily: thymidine kinase  
C:Keywords: phosphotransferase

Query Match 82.8%; Score 24; DB 2; Length 195;  
Best Local Similarity 57.1%; Pred. No. 65;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VMXVAEF 7  
| : |||  
Db 131 LMAIAEF 137

## RESULT 10

B70728  
hypothetical protein RV2558 - *Mycobacterium tuberculosis* (strain H37RV)  
C:Species: *Mycobacterium tuberculosis*  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: B70728  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: B70728  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-236 <COL>  
A:Cross-references: GB:277250; GB:AL123456; NID:g3261617; PIDN:CAB01046.1; PID:e25533  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV2558

Query Match 82.8%; Score 24; DB 2; Length 236;  
Best Local Similarity 71.4%; Pred. No. 79;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VMXVAEF 7  
| : |||  
Db 217 VLDVAEF 223

## RESULT 11

T18315  
hypothetical protein L7610.5 - *Leishmania major*  
C:Species: *Leishmania major*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T18315  
R;Oliver, K.; Murphy, L.; Quail, M.; Lawson, D.; Harris, D.; Rajandream, M.; Ivens, A.;  
submitted to the EMBL Data Library, May 1999  
A;Reference number: Z18876  
A;Accession: T18315  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-248 <OLI>  
A;Cross-references: EMBL:AL034356; NID:el371878; PID:el371560; PIDN:CAA22241.1  
C;Genetics:  
A;Note: L7610.5

Query Match 82.8%; Score 24; DB 2; Length 248;  
Best Local Similarity 71.4%; Pred. No. 83;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
: |||||  
Db 196 LMRVAEF 202

## RESULT 12

G81693  
serine/threonine protein phosphatase, probable TC0530 [imported] - Chlamydia muridarum  
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000  
C;Accession: G81693  
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Bodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A;Reference number: A81500; MUID:20150255; PMID:10684935  
A;Accession: G81693  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-249 <TE>  
A;Cross-references: GB:AE002321; GB:AE002160; NID:g7190560; PIDN:AAF39372.1; PID:g719057  
A;Experimental source: strain Nigg (MoPn)  
C;Genetics:  
A;Gene: TC0530  
C;Superfamily: conserved hypothetical protein yloO; conserved hypothetical protein yloO

Query Match 82.8%; Score 24; DB 2; Length 249;  
Best Local Similarity 71.4%; Pred. No. 84;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
: |||||  
Db 1 MMIVAEF 7

## RESULT 13

F95103  
6-phosphofructokinase [imported] - Streptococcus pneumoniae (strain TIGR4)  
C;Species: Streptococcus pneumoniae  
C;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001  
C;Accession: F95103  
R;Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
son, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A;Reference number: A95000; MUID:21357209; PMID:11463916  
A;Accession: F95103  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-335 <KUR>  
A;Cross-references: GB:AE005672; PIDN:AAK75023.1; PID:gl4972371; GSPDB:GN00164; TIGR:SP4  
A;Experimental source: strain TIGR4  
C;Genetics:  
A;Gene: SP0896  
C;Superfamily: 6-phosphofructokinase; 6-phosphofructokinase 1 homology

Query Match 82.8%; Score 24; DB 2; Length 335;  
Best Local Similarity 71.4%; Pred. No. 11e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
: |||||  
Db 224 VMSAAEF 230

## RESULT 14

D97371  
6-phosphofructokinase (EC 2.7.1.11) [imported] - Streptococcus pneumoniae (strain  
C;Species: Streptococcus pneumoniae  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
C;Accession: D97371  
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff,  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhre  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskuna  
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A;Reference number: A97872; MUID:21429245; PMID:11544234  
A;Accession: D97371  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-335 <KUR>  
A;Cross-references: GB:AE007317; PIDN:AAK99600.1; PID:gl5458395; GSPDB:GN00174  
C;Genetics:  
A;Gene: pfkA  
C;Superfamily: 6-phosphofructokinase; 6-phosphofructokinase 1 homology  
C;Keywords: phosphotransferase

Query Match 82.8%; Score 24; DB 2; Length 335;  
Best Local Similarity 71.4%; Pred. No. 11e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
: |||||  
Db 224 VMSAAEF 230

## RESULT 15

F97010  
fusion, chorismate mutase and shikimate 5-dehydrogenase [imported] - Clostridium  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C;Accession: F97010  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson,  
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacte  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: F97010  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-367 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK78873.1; PID:gl5023795; GSPDB:GN00168  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC0897

Query Match 82.8%; Score 24; DB 2; Length 367;  
Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
: |||||  
Db 29 VMEVAEY 35

Search completed: April 23, 2003, 12:36:26  
Job time : 9.44828 secs







GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:33:40 ; Search time 4.58621 Seconds  
(without alignments)  
63.306 Million cell updates/sec

Title: US-09-723-722a-78

Perfect score: 29

Sequence: 1 VMXVAEF 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	28	96.6	1 POLG_LANVT	P29837 1 genome po
2	26	89.7	1 DEOC_MYCGE	P47296 mycoplasma
3	26	89.7	1 DEOC_MYCPN	P09924 mycoplasma
4	25	86.2	1 YTBQ_BACSU	P53560 bacillus su
5	25	86.2	1 ENFD_SCHPO	P87111 s probabile
6	24	82.8	1 RL13_HALMA	P29198 haloarcula
7	24	82.8	1 RAP_TAROF	O49065 taraxacum o
8	24	82.8	1 YP58_MYCTU	Q50740 mycobacteri
9	24	82.8	1 K6PF_STRPN	P83297 drosophila
10	24	82.8	1 G64F_DROME	P97rc6 streptococc
11	24	82.8	1 TP6B_AERPE	O9ye64 aeropyrum p
12	24	82.8	1 SYL_BACSU	Q43477 bacillus su
13	24	82.8	1 AMPN_HAECC	Q10737 haemophilus
14	24	82.8	1 AT9B_MOUSE	P98195 mus musculus
15	23	79.3	1 RISB_ARCFU	O28152 archaeoglob
16	23	79.3	1 AROK_LACLA	Q9ceul lactococcus
17	23	79.3	1 PWM_YEAST	P07283 saccharomyc
18	23	79.3	1 NIJM_MYTED	Q00860 mytilus edu
19	23	79.3	1 AMPM_ARCFU	O28438 archaeoglob
20	23	79.3	1 RU22_DROME	P50887 drosophila
21	23	79.3	1 CYCR_ROSDE	P26278 roseobacter
22	23	79.3	1 NTP1_MFXVL	Q9q814 myxoma viru
23	23	79.3	1 NTP1_SFVKA	O9q822 Shope fibro
24	23	79.3	1 BPR2_HUMAN	Q01742 homo sapien
25	23	79.3	1 GYRB_HELPY	Q92133 helicobacte
26	23	79.3	1 GYRB_HELPY	P55992 helicobacte
27	23	79.3	1 FGR3_MOUSE	P61851 mus musculus
28	23	79.3	1 FGR4_HUMAN	P22455 homo sapien
29	23	79.3	1 CEK2_CHICK	P18460 gallus gall
30	23	79.3	1 FGR3_HUMAN	P22607 homo sapien
31	23	79.3	1 FGR4_MOUSE	Q03142 mus musculus
32	23	79.3	1 FGR2_XENLA	Q03364 xenopus lae
33	23	79.3	1 FGR1_CHICK	P21804 gallus gall

34 23 79.3 821 1 FGR2\_HUMAN P21802 homo sapien  
35 23 79.3 821 1 FGR2\_MOUSE P21803 mus musculus  
36 23 79.3 822 1 FGR1\_HUMAN P11362 homo sapien  
37 23 79.3 822 1 FGR1\_MOUSE P16092 mus musculus  
38 23 79.3 822 1 FGR1\_RAT Q04589 rattus norv  
39 23 79.3 823 1 CER3\_CHICK P18461 gallus gall  
40 23 79.3 852 1 VTER\_HSV1 Q00140 ictalurid h  
41 23 79.3 980 1 POLG\_LIV P22338 louping ill  
42 23 79.3 1002 1 EPB5\_CHICK Q07497 gallus gall  
43 23 79.3 1052 1 FGR2\_DROME Q09147 drosophila  
44 23 79.3 1914 1 STCK\_EMENI Q00706 emericeella  
45 23 79.3 3412 1 POLG\_TBEVS P07720 t genome po

#### ALIGNMENTS

##### RESULT 1

POLG\_LANVT  
ID POLG\_LANVT STANDARD; PRT; 3414 AA.  
AC P29837;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (Core protein);  
DE Envelope protein PM; Matrix protein (Envelope protein M); Major  
DE envelope protein E; Nonstructural protein NS1; Nonstructural protein  
DE NS2A; Nonstructural protein NS2B; Helicase/protease (EC 3.4.21.98)  
DE (NS3); Nonstructural protein NS4A; Nonstructural protein NS4B; RNA-  
DE directed RNA polymerase (EC 2.7.7.48) (NS5)].  
OS Langat virus (strain TP21).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Flavivirus  
OX NCBI\_TaxID=31638;  
RN [1]  
RP SEQUENCE OF 1-776 FROM N.A.  
RX MEDLINE=92074260; PubMed=1720591;  
RA Mandl C.W., Iacono-Connors L., Wallner G., Holzmann H., Kunz C.,  
RA Heinz F.X.;  
RT "Sequence of the genes encoding the structural proteins of the low-  
RT virulence tick-borne flaviviruses Langat TP21 and Yelantsev";  
RL Virology 185:891-895(1991).  
RN [2]  
RX SEQUENCE OF 777-3414 FROM N.A.  
RX MEDLINE=92263794; PubMed=1316684;  
RA Iacono-Connors L.C., Schmaljohn C.S.;  
RT "Cloning and sequence analysis of the genes encoding the  
RT nonstructural proteins of Langat virus and comparative analysis with  
RT other flaviviruses";  
RL Virology 188:875-880(1992).  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC [RNA](N).  
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M73835; AAA02740.1; ALT\_TERM.  
CC EMBL; S35365; AAB22165.1; -.  
CC PIR; A41704; A41704.  
CC PIR; A42545; A42545.  
CC HSSP; P14336; 1SVB.  
CC -----

```
DR MEROPS; S07.001; -
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR000069; Flavi_M.
DR InterPro: IPR001157; Flavi_NSI.
DR InterPro: IPR000752; Flavi_NS2A.
DR InterPro: IPR000487; Flavi_NS2B.
DR InterPro: IPR000404; Flavi_NS4A.
DR InterPro: IPR001528; Flavi_NS4B.
DR InterPro: IPR000208; Flavi_NS5.
DR InterPro: IPR001122; Flavi_capsidC.
DR InterPro: IPR000336; Flavi_glycoprote.
DR InterPro: IPR001850; Flavi_helicase.
DR InterPro: IPR002535; Flavi_propep.
DR InterPro: IPR002877; FtsJ.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00869; Flavi_glycoprot; 1.
DR Pfam: PF00948; Flavi_NSI; 1.
DR Pfam: PF00949; Flavi_helicase; 1.
DR Pfam: PF00972; Flavi_NS5; 1.
DR Pfam: PF01002; Flavi_NS2B; 1.
DR Pfam: PF01003; Flavi_NS2B; 1.
DR Pfam: PF01004; Flavi_M; 1.
DR Pfam: PF01005; Flavi_NS2A; 1.
DR Pfam: PF01349; Flavi_NS4B; 1.
DR Pfam: PF01350; Flavi_NS4A; 1.
DR Pfam: PF01570; Flavi_propep; 1.
DR Pfam: PF01728; FtsJ; 1.
DR Pfam: PF02832; Flavi_glycop_C; 1.
DR ProDom: PD001496; Flavi_NSI; 1.
DR ProDom: PD001556; Flavi_glycoprote; 1.
DR SMART; SM00490; HELIC; 1.
KW Polypeptide; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;
KW ATP-binding; Transmembrane; Nonstructural protein.
FT INIT_MET 1 1
  CHAIN 1 112
  CHAIN 113 205
  CHAIN 206 280
  CHAIN 281 776
  CHAIN 777 1128
  CHAIN 1129 1358
  CHAIN 1359 1489
  CHAIN 1490 2110
  CHAIN 2111 2259
  CHAIN 2260 2511
  CHAIN 2512 3414
  NP_BIND 1688 1695
  SITE 1779 1782
  TRANSMEM 103 119
  TRANSMEM 262 278
  TRANSMEM 728 744
  TRANSMEM 758 774
  DISULFID 283 310
  DISULFID 340 396
  DISULFID 354 385
  DISULFID 372 401
  DISULFID 466 570
  DISULFID 587 618
  CARBOHYD 144 144
  CARBOHYD 434 434
  SEQUENCE 3414 AA; 378017 MW; 59CB7E95DD70D82E CRC64;
Query Match 96.6%; Score 28; DB 1; Length 3414;
Best Local Similarity 85.7%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 VMXVAEF 7
  || |||||
Db 929 VMTVAEF 935
```

```
RESULT 2
DEOC_MYCGE
ID DEOC_MYCGE STANDARD; PRT; 223 AA.
AC P47296;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyriboaldolase)
DE (Deoxyriboaldolase).
GN DEOC OR MG050.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bitt K.F., Hu P.-C., Lucier J.C.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
CC -!- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate - D-
CC glyceraldehyde 3-phosphate + acetaldehyde.
CC -!- PATHWAY: Nucleotide and deoxyribonucleotide catabolism.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE DEOC/PBAB FAMILY OF ALDOLASES.
CC DEOC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; U39684; AAC71266.1; -.
CC TIGR; MG050; -.
CC InterPro: IPR002915; Deoc.
CC InterPro: IPR003009; FMN_enzyme.
CC Pfam; PF01791; Deoc; 1.
CC Lyase; Schiff base; Complete proteome.
KW BINDING 152 152
FT BINDING 152 152
FT SEQUENCE 223 AA; 24675 MW; 332430231CE99DB0 CRC64;
Query Match 89.7%; Score 26; DB 1; Length 223;
Best Local Similarity 71.4%; Pred. No. 9;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 VMXVAEF 7
  || |||||
Db 93 VMNIAEF 99

RESULT 3
DEOC_MYCPCN
ID DEOC_MYCPCN STANDARD; PRT; 224 AA.
AC P09924;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyriboaldolase)
DE (Deoxyriboaldolase).
GN DEOC OR MPN063 OR MP091.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
```

RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29342 / M129;  
 RX MEDLINE=89128453; PubMed=2492658;  
 RA Loechel S., Inamine J.M., Hu P.-C.;  
 RT "Nucleotide sequence of the deoC gene of Mycoplasma pneumoniae.";  
 RL Nucleic Acids Res. 17:801-801(1989).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29342 / M129;  
 RX MEDLINE=97105885; PubMed=8948633;  
 RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,  
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.";  
 RL Nucleic Acids Res. 24:4420-4449(1996).  
 CC -1- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate = D-  
 CC glyceraldehyde 3-phosphate + acetaldehyde.  
 CC -1- PATHWAY: Nucleotide and deoxyribonucleotide catabolism.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE DEOC/FBAP FAMILY OF ALDOLASES.  
 CC DEOC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X13544; CAA31897.1; -  
 CC EMBL: AE000011; AAB95739.1; -  
 CC PIR: S02216; S02216.  
 CC InterPro: IPR002915; Deoc.  
 CC Pfam: PF01791; Deoc; 1.  
 CC Lyase; Schiff base; Complete proteome.  
 KW BINDING 152 152 SCHIFF-BASE (BY SIMILARITY).  
 FT SEQUENCE 224 AA; 24878 MW; 73C3E4932E78B1F7 CRC64;  
 SQ  
 Query Match 89.7%; Score 26; DB 1; Length 224;  
 Best Local Similarity 71.4%; Pred. No. 9;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VMXVAEF 7  
 DB 93 VMNIAEF 99  
 ID YTBQ\_BACSU STANDARD; PRT; 253 AA.  
 AC P53560; O34914;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein ytbQ.  
 GN YTBQ.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96312354; PubMed=8763940;  
 RA Bower S., Perkins J.B., Vocum R.R., Howitt C.L., Rahaim P.,  
 RA Pero J.;  
 RT "Cloning, sequencing, and characterization of the Bacillus subtilis  
 RT biotin biosynthetic operon.";  
 RL J. Bacteriol. 178:4122-4130(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]

RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Chim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardiniois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis.";  
 RL Nature 390:249-256(1997).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U51868; AAB17463.1; -  
 CC EMBL: AF008220; AAC00267.1; ALT\_INIT.  
 CC EMBL: Z99119; CAB14996.1; ALT\_INIT.  
 CC Subtilist; BG11787; ytbQ.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 253 AA; 28233 MW; 572CC4681F04C650 CRC64;  
 Query Match 86.2%; Score 25; DB 1; Length 253;  
 Best Local Similarity 57.1%; Pred. No. 19;  
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 VMXVAEF 7  
 DB 75 IMDIAEF 81  
 ID ETFD\_SCHPO STANDARD; PRT; 632 AA.  
 AC P87111;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable electron transfer flavoprotein-ubiquinone oxidoreductase,  
 DE mitochondrial precursor (EC 1.5.5.1) (ETf-QUO) (ETf-ubiquinone  
 DE oxidoreductase) (ETf dehydrogenase) (Electron-transferring-  
 DE flavoprotein dehydrogenase).  
 GN SPAC2068.04C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gilliam R., Rayandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Money P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzyp K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomycetes pombe";  
 RL Nature 415:871-880(2002).  
 CC -!- FUNCTION: ACCEPTS ELECTRONS FROM ETF AND REDUCES UBIQUINONE (BY  
 CC SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: Reduced ETF + ubiquinone -> ETF + ubiquinol.  
 CC -!- COFACTOR: FAD AND A 4FE-4S CLUSTER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane (By  
 CC similarity).  
 CC -!- SIMILARITY: BELONGS TO THE ETF-OO / FIXC FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: Z95334; CAB08598.1; -;  
 DR Oxidoreductase; Electron transport; Flavoprotein; FAD; Iron-sulfur;  
 KW 4Fe-4S; Mitochondrion; Transit peptide; Ubiquinone.  
 FT TRANSIT 1 ?  
 FT CHAIN ? 632  
 FT PROBABLE ELECTRON TRANSFER FLAVOPROTEIN-  
 FT UBIQUINONE OXIDOREDUCTASE.  
 FT FAD (ADP PART) (POTENTIAL).  
 FT NP\_BIND 93 107  
 FT IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 575 575  
 FT IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 601 601  
 FT IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 604 604  
 FT IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 607 607  
 FT - IRON-SULFUR (4FE-4S) (POTENTIAL).  
 SQ SEQUENCE 632 AA; 69472 MW; 1B0P22374E33771B CRC64;  
 Query Match 86.2%; Score 25; DB 1; Length 632;  
 Best Local Similarity 71.4%; Pred. No. 46;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 VNXVAEF 7  
 Db 193 VMSLAEF 199  
 RESULT 6  
 RL13\_HALMA

RL13\_HALMA STANDARD; PRT; 145 AA.  
 P29198;  
 AC 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 50S ribosomal protein L13p (Hmali3).  
 GN RPL13P.  
 OS Haloarcula marismortui (Haloaracterium marismortui).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Haloarcula.  
 OX NCBI\_TaxID=2238;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=92105119; PubMed=1840597;  
 RX Kroemer W.J., Arndt E.;  
 RA "Halobacterial S9 operon. Three ribosomal protein genes are  
 RT cotranscribed with genes encoding a tRNA(Leu), the endonuclease, and a  
 RT putative membrane protein in the archaeobacterium Haloarcula  
 RT (Haloaracterium marismortui";  
 RL J. Biol. Chem. 266:24573-24579(1991).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 RC STRAIN=ATCC 43049;  
 RX MEDLINE=20396344; PubMed=10937989;  
 RA Ban N., Nissen P., Hansen J., Moore P.B., Steitz T.A.;  
 RT "The complete atomic structure of the large ribosomal subunit at 2.4  
 RT A resolution";  
 RL Science 289:905-920(2000).  
 CC -!- SIMILARITY: BELONGS TO THE L13P FAMILY OF RIBOSOMAL PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M76567; AAA73097.1; -;  
 DR PIR: B41715; B41715.  
 DR PDB: 1FFK; 14-AUG-00.  
 DR InterPro: IPR001074; Ribosomal\_L13.  
 DR Pfam: PF00572; Ribosomal\_L13; 1.  
 DR TIGRFAMs: TIGR01077; L13\_A\_E; 1.  
 DR PROSITE: PS00783; RIBOSOMAL\_L13; 1.  
 KW Ribosomal protein; 3D-structure.  
 SQ SEQUENCE 145 AA; 16228 MW; 069CE566662AE3BC CRC64;  
 Query Match 82.8%; Score 24; DB 1; Length 145;  
 Best Local Similarity 83.3%; Pred. No. 21;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 MXVAEF 7  
 Db 1 MSVAEF 6  
 RESULT 7  
 RAP\_TAROF  
 ID RAP\_TAROF STANDARD; PRT; 157 AA.  
 AC O49065;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Root allergen protein (RAP).  
 OS Taraxacum officinale (Common dandelion).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;  
 OC Taraxacum.  
 OX NCBI\_TaxID=50225;  
 RN [1]  
 RP SEQUENCE FROM N.A.

CC TISSUE=Root;  
 RA Xu X.-Y., Bewley J.D., Greenwood J.S.;  
 RL Submitted (DSC-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE BETVI FAMILY OF PATHOGENESIS-RELATED  
 CC PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF036931; AAB92255.1; -  
 DR HSSP; O24248; 1809  
 DR InterPro; IPR000916; Bet\_v\_I.  
 DR Pfam; PF00407; Bet\_v\_I; 1.  
 DR PRINTS; PR00634; BETALLERGEN.  
 DR PROSITE; PS00451; PATHOGENESIS\_BETVI; 1.  
 KW Allergen.  
 SQ SEQUENCE 157 AA; 17040 MW; 5892AB8593A8A7E0 CRC64;  
 -----  
 Query Match 82.8%; Score 24; DB 1; Length 157;  
 Best Local Similarity 83.3%; Pred. No. 22;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 2 MXVAEF 7  
 DB 1 MVADEF 6  
 I: |||||  
 -----  
 RESULT 8  
 YP58\_MYCTU  
 ID YP58\_MYCTU STANDARD; PRT; 236 AA.  
 AC Q50740;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein RV2558.  
 GN RV2558 OR MT2635 OR WTCY9C4.10C.  
 OS Mycobacterium tuberculosis.  
 CC Actinobacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: STRONG, TO M.TUBERCULOSIS RV2557.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z77250; CAB01046.1; -  
 DR EMBL; AE007098; AAK46947.1; -  
 DR TIGR; MT2635; -  
 DR TubercuList; RV2558; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 236 AA; 25718 MW; 13E3B049D8F79C6B CRC64;  
 -----  
 Query Match 82.8%; Score 24; DB 1; Length 236;  
 Best Local Similarity 71.4%; Pred. No. 33;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 VMXVAEF 7  
 DB 217 VLDVAEF 223  
 I: |||||  
 -----  
 RESULT 9  
 K6PF\_STRPN  
 ID K6PF\_STRPN STANDARD; PRT; 335 AA.  
 AC Q97RC6;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 6-phosphofructokinase (EC 2.7.1.11) (Phosphofructokinase)  
 DE (Phosphohexokinase).  
 GN PFKA OR SP0896  
 OS Streptococcus pneumoniae.  
 CC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 CC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Tettelin H., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
 RA Peterson S., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L., E.K.,  
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus  
 RT pneumoniae."  
 RL Science 293:498-506(2001).  
 CC -!- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate - ADP + D-  
 CC fructose 1,6-bisphosphate.  
 CC -!- PATHWAY: Key control step of glycolysis.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE007395; AAK75023.1; -  
 DR TIGR; SP0896; -  
 DR InterPro; IPR000023; Ppfrckinase.  
 DR Pfam; PF00365; PFK; 1.  
 DR PRINTS; PR00476; PPFCKINASE.  
 DR PRODOM; PD000707; Ppfrckinase; 1.

DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.  
 KW Kinase; Transferase; Glycolysis; Complete proteome.  
 SQ SEQUENCE 335 AA; 35174 MW; 2DFFBAACC58CAF66 CRC64;

Query Match 82.8%; Score 24; DB 1; Length 335;  
 Best Local Similarity 71.4%; Pred. No. 47;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VNXVAEF 7

DB 224 VNSAAEF 230

# RESULT 10

G64F\_DROME STANDARD; PRT; 449 AA.

AC P83297; Q3VZJ5;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DE Putative gustatory receptor 64f.

GN GR64F OR CG14988.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkeley;

RA MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cavley S., Dahke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fogle C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

RN [2]

RP IDENTIFICATION.

RX MEDLINE=21407712; PubMed=11516643;

RA Dunipace L., Meister S., McNealy C., Amrein H.;

RT "Spatially restricted expression of candidate taste receptors in the  
 RT Drosophila gustatory system.";  
 RL Curr. Biol. 11:822-835(2001).

RN [3]

RP CONCEPTUAL TRANSLATION.

RA Robertson H.;

RL Unpublished observations (NOV-2001).

CC -!- FUNCTION: Probable role in the gustatory response.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -!- SIMILARITY: BELONGS TO FAMILY DR-TR OF G-PROTEIN COUPLED

CC RECEPTORS. SUBFAMILY II.

CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous

CC gene model prediction.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/))  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL; AE003480; AAF47826.1; ALT\_SEQ.

DR FlyBase; FBgn0045475; Gr64f.

DR InterPro; IPR004117; 7tm\_6.

DR Pfam; PF02949; 7tm\_6; 1.

KW Hypothetical protein; Receptor; G-protein coupled receptor;

KW Transmembrane; Multigene family.

FT DOMAIN 1 97 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 98 118 1 (POTENTIAL).

FT DOMAIN 119 126 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 127 147 2 (POTENTIAL).

FT DOMAIN 148 179 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 180 200 3 (POTENTIAL).

FT DOMAIN 201 245 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 246 266 4 (POTENTIAL).

FT DOMAIN 267 310 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 311 331 5 (POTENTIAL).

FT DOMAIN 332 354 6 (POTENTIAL).

FT TRANSMEM 355 415 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 416 436 7 (POTENTIAL).

FT TRANSMEM 437 449 CYTOPLASMIC (POTENTIAL).

SQ SEQUENCE 449 AA; 52190 MW; 109D965C6E03BA44 CRC64;

Query Match 82.8%; Score 24; DB 1; Length 449;

Best Local Similarity 71.4%; Pred. No. 62;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VNXVAEF 7

DB 64 VLLVAEF 70

# RESULT 11

TP6B\_AERPE STANDARD; PRT; 565 AA.

ID TP6B\_AERPE

AC Q9VE64;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Type II DNA topoisomerase VI subunit B (EC 5.99.1.3).

GN TOP6B OR APE0706.

OS Aeropyrum pernix.

OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;

OC Desulfurococaceae; Aeropyrum.

OX NCBI\_TaxID=56636;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K1.

RX MEDLINE=99310339; PubMed=10382966;

RA Kwarabiyasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,

RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,

RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,  
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
RT crenarchaeon, Aeropyrum pernix K1.";  
RA DNA Res. 6:83-101(1999).  
CC -1- FUNCTION: RELAXES BOTH POSITIVE AND NEGATIVE SUPERURNS AND  
CC EXHIBITS A STRONG DECATENASE ACTIVITY. THE B SUBUNIT BINDS ATP (BY  
CC SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining  
CC of double-stranded DNA.  
CC -1- SUBUNIT: HETEROTETRAMER OF TWO SUBUNITS A AND TWO SUBUNITS B (BY  
CC SIMILARITY).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AP000060; BAA79682.1; -  
CC InterPro; IPR003594; ATPbind\_ATPase.  
CC InterPro; IPR004359; HIS\_KIN\_sig.  
CC Pfam; PF02518; HATPase\_C; 1.  
CC SMART; SM00387; HATPase\_c; 1.  
CC TIGRfams; TIGR01052; top6b; 1.  
CC Isomerase; Topoisomerase; DNA-binding; ATP-binding; Complete proteome.  
CC KW SEQUENCE 565 AA; 64096 MW; BAA3DFCC5E3D508 CRC64;  
CC -----  
CC Query Match 82.8%; Score 24; DB 1; Length 565;  
CC Best Local Similarity 83.3%; Pred. No. 77;  
CC Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
CC -----  
CC QY 2 MXVAEF 7  
CC | |||||  
CC Db 28 MSVAEF 33  
CC -----  
CC RESULT 12  
CC SVI\_BACSU STANDARD; PRT; 921 AA.  
CC AC Q45477; P71022; O31730;  
CC DT 01-NOV-1997 (Rel. 35, Created)  
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)  
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
CC DE Isoleucyl-tRNA synthetase (EC 6.1.1.5) (Isoleucine--tRNA ligase)  
CC GN (Ilers).  
CC OS Bacillus subtilis.  
CC OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
CC OX NCBI\_TaxID=1423;  
CC [1]  
CC RN SEQUENCE FROM N.A.  
CC RC STRAIN=168;  
CC RX MEDLINE=38044033; PubMed=9384377;  
CC RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
CC Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
CC Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
CC Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
CC Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
CC Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
CC Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
CC Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,  
CC Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
CC Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
CC Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
CC Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
CC Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
CC Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
CC Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
RL [2]  
RN SEQUENCE OF 1-121 FROM N.A.  
RP STRAIN=168;  
RC Stewart G.C., Cha J.H.;  
RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RL [3]  
RP SEQUENCE OF 733-921 FROM N.A.  
RC STRAIN=168;  
RA Pragai Z., Tjalsma H., Bolhuis A., van Dijk J.M., Venema G.,  
RA Bron S.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: ATP + L-isoleucine + tRNA(Ile) = AMP +  
CC diphosphate + L-isoleucyl-tRNA(Ile).  
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z99112; CAB13417.1; -  
CC EMBL; U60901; AAB49280.1; -  
CC EMBL; U48870; AAB57764.1; -  
CC HSP; P41972; IQ03.  
CC DR Subtilisin; BGI1792; Iles.  
CC DR InterPro; IPR002300; tRNA-synt\_1a.  
CC DR InterPro; IPR001412; tRNA-synt\_1.  
CC DR InterPro; IPR002301; tRNA-synt\_ile.  
CC DR Pfam; PF00133; tRNA-synt\_1; 1.  
CC DR PRINTS; PR00984; TRNASYNTHILE.  
CC DR TIGRfams; TIGR00392; Iles; 1.  
CC DR PROSITE; PS00178; AA-TRNA-LIGASE\_I; 1.  
CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
CC Metal-binding; Zinc; Complete proteome.  
CC FT SITE 57 67 "HIGH" REGION.  
CC FT SITE 594 598 "KMSKS" REGION.  
CC FT BINDING 597 597 ATP (BY SIMILARITY).  
CC FT CONFLICT 744 744 S -> F (IN REF. 3).  
CC SQ SEQUENCE 921 AA; 104784 MW; 011F31E5F7460D43 CRC64;  
CC -----  
CC Query Match 82.8%; Score 24; DB 1; Length 921;  
CC Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
CC Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
CC -----  
CC QY 2 MXVAEF 7  
CC | |||||  
CC Db 116 MSVAEF 121  
CC -----  
CC RESULT 13  
CC AMPN\_HAE00



ID AC AMPL\_HAECO STANDARD; PRT; 971 AA.  
 DT Q10737;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Aminopectidase N (EC 3.4.11.2) (Microsomal aminopectidase) (Membrane  
 DE glycoprotein H11).  
 OS Haemophilus contortus (Barber pole worm).  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
 OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.  
 ON NCBI\_TaxID=6289;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97273974; PubMed=9128148;  
 RA Smith T.S., Graham M., Munn E.A., Newton S.E., Knox D.P.,  
 RA Coadwell W.J., McMichael-Phillips D., Smith H., Smith W.D.,  
 RA Oliver J.J.;  
 RT "Cloning and characterization of a microsomal aminopectidase from the  
 RT intestine of the nematode Haemonchus contortus";  
 RL Biochim. Biophys. Acta 1338:295-306(1997).  
 CC -|- CATALYTIC ACTIVITY: Release of an N-terminal amino acid,  
 CC preferably a neutral or hydrophobic one, from a polypeptide.  
 CC Aminoacyl-arylamides are poor substrates.  
 CC -|- COFACTOR: BINDS 1 ZINC ION.  
 CC -|- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).  
 CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X94187; CAA63897.1;  
 DR MEROPS; M01.015;  
 DR InterPro; IPR001930; Ala\_peptase.  
 DR InterPro; IPR000130; Zn\_Mtpeptidase.  
 DR Pfam; PF01433; Peptidase\_M1; 1.  
 DR PRINTS; PR00756; ALADIPTASE.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Aminopectidase; Zinc; Glycoprotein;  
 KW Transmembrane; Signal-anchor.  
 FT INIT\_MET 0  
 FT DOMAIN 1 16  
 FT TRANSMEM 17 38  
 FT  
 FT DOMAIN 39 971  
 FT METAL 378 378  
 FT ACT\_SITE 379 379  
 FT METAL 382 382  
 FT METAL 401 401  
 FT ACT\_SITE 465 465  
 FT CARBOHYD 98 98  
 FT CARBOHYD 226 226  
 FT CARBOHYD 548 548  
 FT CARBOHYD 857 857  
 SQ SEQUENCE 971 AA; 110542 MW; 95C6A92B5CCA227C CRC64;  
  
 Query Match 82.8%; Score 24; DB 1; Length 971;  
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 1 VMXVAEF 7  
 I: | | | |  
 Db 541 VISVAEF 547  
  
 RESULT 14  
 AT9B\_MOUSE  
 ID AT9B\_MOUSE  
 AC P98195; STANDARD; PRT; 1095 AA.

DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Potential phospholipid-transporting ATPase IIB (EC 3.6.3.1).  
 GN ATP9B.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ICR; TISSUE=Brain;  
 RX MEDLINE=20473714; PubMed=11015572;  
 RA Hallett M.S., Lawler J.F. Jr., Blackshaw S., Gao L., Nagarajan P.,  
 RA Hacker C., Pyle S., Newman J.T., Nakanishi Y., Ando H., Weinstein D.,  
 RA Williamson P., Schlegel R.A.;  
 RT "Differential expression of putative transbilayer amphipath  
 RT transporters";  
 RL Physiol. Genomics 1:139-150(1999).  
 CC -|- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.  
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -|- TISSUE SPECIFICITY: FOUND IN MOST TISSUES EXCEPT SPLEEN AND  
 CC MUSCLE. MOST ABUNDANT IN TESTIS. ALSO DETECTED IN FETAL TISSUES.  
 CC -|- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2  
 CC ATPASES). SUBFAMILY IV.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF155913; AAF08476.1;  
 DR MGD; MGI-1354757; Atp9b.  
 DR InterPro; IPR001757; ATPase\_E1-E2.  
 DR InterPro; IPR001454; Hignase/hydrlase.  
 DR Pfam; PF00122; E1-E2\_ATPase; 1.  
 DR Pfam; PF00702; Hydrolase; 1.  
 DR PRINTS; PR00119; CATATPASE.  
 DR PROSITE; PS00154; ATPASE\_E1\_E2; 1.  
 KW Hydrolase; transmembrane; Phosphorylation; Magnesium; ATP-binding;  
 KW Multigene family.  
 FT DOMAIN 1 94  
 FT TRANSMEM 95 116  
 FT DOMAIN 117 121  
 FT TRANSMEM 122 144  
 FT DOMAIN 145 328  
 FT TRANSMEM 329 349  
 FT DOMAIN 350 357  
 FT TRANSMEM 358 379  
 FT DOMAIN 380 878  
 FT TRANSMEM 879 899  
 FT DOMAIN 900 911  
 FT TRANSMEM 912 930  
 FT DOMAIN 931 960  
 FT TRANSMEM 961 979  
 FT DOMAIN 980 986  
 FT TRANSMEM 987 1009  
 FT DOMAIN 1010 1015  
 FT TRANSMEM 1016 1036  
 FT DOMAIN 1037 1053  
 FT TRANSMEM 1054 1078  
 FT DOMAIN 1079 1095  
 FT MOD\_RES 416 416  
 FT METAL 822 822  
 FT METAL 826 826  
 SQ SEQUENCE 1095 AA; 123506 MW; 76D37684241D176E CRC64;  
  
 Query Match 82.8%; Score 24; DB 1; Length 1095;  
 Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;



Db 8 MVVAEF 13

Search completed: April 23, 2003, 12:35:45  
Job time : 6.58621 secs

Oy 1 VMXVAEF 7  
: | | | | |  
Db 1016 LMVVAEF 1022

RESULT 15

```
RISB_ARCFU
ID RISB_ARCFU STANDARD: PRT: 143 AA.
AC O28152;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)
DE (Lumazine synthase) (Riboflavin synthase beta chain).
GN RIBH OR RIBE OR AF2128.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -!- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
CC catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
CC ribityl-amino-2,4-(1H,3H)-pyrimidinone and L-3,4-dihydroxy-2-
CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-
CC 2,4-(1H,3H)-pyrimidinone with L-3,4-dihydroxy-2-butanone-4-
CC phosphate yielding 6,7-dimethyl-8-lumazine (BY similarity).
CC -!- CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1-D-ribityl)lumazine -
CC riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
CC -!- PATHWAY: Riboflavin biosynthesis; last step.
CC -!- SIMILARITY: BELONGS TO THE DMRL SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000957; AAB89124.1; -
CC HSSP; P11998; IRVV.
CC TIGR; AF2128; -
CC InterPro; IPR002180; DMRL_synthase.
CC Pfam; PF00885; DMRL_synthase; 1.
CC ProDom; PD03664; DMRL_synthase; 1.
CC TIGRFAMs; TIGR00114; ribH; 1.
CC Riboflavin biosynthesis; Transferase; Complete proteome.
CC SEQUENCE 143 AA; 15645 MW; BBA3141206905410 CRC64;
Query Match 79.3%; Score 23; DB 1; Length 143;
Best Local Similarity 83.3%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Oy 2 MXVAEF 7



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:36:34 ; Search time 14.9655 seconds  
(without alignments)  
37.480 Million cell updates/sec

Title: US-09-723-722A-78

Perfect score: 29  
Sequence: 1 VMXVAEF 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	93.1	68	10	US-09-393-634-80
2	24	82.8	145	9	US-10-013-379-11
3	24	82.8	335	10	US-09-815-242-13346
4	24	82.8	409	9	US-10-081-816-14
5	24	82.8	1084	9	US-10-024-623-23
6	24	82.8	1095	9	US-10-024-623-26
7	24	82.8	1095	9	US-10-024-623-36
8	23	79.3	54	9	US-09-158-722-18
9	23	79.3	54	9	US-09-158-722-46
10	23	79.3	54	9	US-09-158-722-47
11	23	79.3	90	9	US-09-510-332-40
12	23	79.3	90	10	US-09-393-634-76
13	23	79.3	150	10	US-09-815-242-4899
14	23	79.3	237	9	US-09-738-626-4458
15	23	79.3	280	9	US-09-738-626-4505
16	23	79.3	310	9	US-09-939-833-7
17	23	79.3	310	10	US-09-939-754-7
18	23	79.3	310	10	US-09-939-832-7
19	23	79.3	314	9	US-09-510-332-44

Sequence 71, Appl  
Sequence 5133, Ap  
Sequence 5303, Ap  
Sequence 12388, A  
Sequence 12820, A  
Sequence 714, App  
Sequence 11330, A  
Sequence 33, Appl  
Sequence 2, Appl  
Sequence 3, Appl  
Sequence 9, Appl  
Sequence 21, Appl  
Sequence 10627, A  
Sequence 19, Appl  
Sequence 110, App  
Sequence 6922, Ap  
Sequence 257, App  
Sequence 6, Appl  
Sequence 216, App  
Sequence 375, App  
Sequence 131, App  
Sequence 24, Appl  
Sequence 7, Appl  
Sequence 13274, A  
Sequence 13532, A

314 9 US-09-510-332-71  
387 10 US-09-815-242-5133  
421 10 US-09-815-242-5303  
421 10 US-09-815-242-12388  
421 10 US-09-815-242-12820  
764 10 US-09-925-302-714  
773 10 US-09-815-242-11330  
802 9 US-10-011-548-33  
802 10 US-09-758-386-3  
822 9 US-09-757-415A-2  
838 9 US-09-927-315-9  
839 9 US-09-897-427A-4  
839 9 US-10-035-045-21  
1208 10 US-09-815-242-10627  
61 9 US-09-978-303-19  
191 9 US-09-860-670-110  
347 9 US-09-738-626-6922  
543 9 US-10-108-605-257  
664 12 US-10-067-385-6  
1007 10 US-09-765-272-216  
90 10 US-09-731-872-375  
144 10 US-09-811-284-131  
149 9 US-09-851-138-24  
160 9 US-09-931-457A-7  
171 10 US-09-815-242-13274  
171 10 US-09-815-242-13532

#### ALIGNMENTS

RESULT 1  
US-09-393-634-80  
; Sequence 80, Application US/09393634  
; Patent No. US20020051997A1  
; GENERAL INFORMATION:  
; APPLICANT: Zuker, Charles S.  
; APPLICANT: Adler, Jon Elliot  
; APPLICANT: Ryba, Nick  
; APPLICANT: Mueller, Ken  
; APPLICANT: Hoon, Mark  
; APPLICANT: The Regents of the University of California  
; APPLICANT: The Government of the United States of America  
; APPLICANT: as represented by the Secretary of the  
; APPLICANT: Department of Health and Human Services  
; TITLE OF INVENTION: SF, a No. US20020051997A1e1 Family of Taste Receptors  
; FILE REFERENCE: 02307E-0980000US  
; CURRENT APPLICATION NUMBER: US/09/393,634  
; CURRENT FILING DATE: 1999-09-10  
; NUMBER OF SEQ ID NOS: 92  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 80  
; LENGTH: 68  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human GR24  
US-09-393-634-80

Query Match 93.1%; Score 27; DB 10; Length 68;  
Best Local Similarity 71.4%; Pred. No. 3.6;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
; : | | | |  
Db 11 IMAVAEF 17

RESULT 2  
US-10-013-379-11  
; Sequence 11, Application US/10013379  
; Publication No. US20020188108A1  
; GENERAL INFORMATION:

APPLICANT: The Regents of the University of California  
APPLICANT: Cate, Jamie H.  
APPLICANT: NO. US20020188108A11er, Harry F.  
APPLICANT: Yusupov, Marat M.  
APPLICANT: Yusupova, Guinara ZH  
APPLICANT: Baucom, Albion  
APPLICANT: Lancaster, Laura  
APPLICANT: Dallas, Anne  
TITLE OF INVENTION: X-RAY CRYSTAL STRUCTURES OF FUNCTIONAL  
TITLE OF INVENTION: COMPLEXES OF THE BACTERIAL RIBOSOME CONTAINING TRANSFER RNA  
TITLE OF INVENTION: AND MODEL MESSENGER RNAs  
FILE REFERENCE: 19629-7010  
CURRENT APPLICATION NUMBER: US/10/013,379  
PRIOR FILING DATE: 2002-03-18  
PRIOR APPLICATION NUMBER: US 60/254,603  
PRIOR FILING DATE: 2000-12-09  
PRIOR APPLICATION NUMBER: US 60/278,013  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: US 60/294,394  
PRIOR FILING DATE: 2001-05-30  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 11  
LENGTH: 145  
TYPE: PRT  
ORGANISM: Haloarcula marismortui  
FEATURE:  
OTHER INFORMATION: 50S ribosomal protein L13  
OTHER INFORMATION: 191yM  
US-10-013-379-11

Query Match 82.8%; Score 24; DB 9; Length 145;  
Best Local Similarity 83.3%; Pred. No. 48;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 MVXAEF 7  
| | | | |  
Db 1 MSVAEF 6

RESULT 3  
US-09-815-242-13346  
Sequence 13346, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 13346  
LENGTH: 335  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-815-242-13346

Query Match 82.8%; Score 24; DB 10; Length 335;  
Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
| | | | |  
Db 224 VMSAAEF 230

RESULT 4  
US-10-081-816-14  
Sequence 14, Application US/10081816  
Publication No. US20030045472A1  
GENERAL INFORMATION:  
APPLICANT: Axel, Richard  
APPLICANT: Scott, Kristin  
TITLE OF INVENTION: Chemosensory Gene Family Encoding Gustatory And Olfactory  
TITLE OF INVENTION: Uses Thereof  
FILE REFERENCE: 0575/64019-A/JPW/ADM  
CURRENT APPLICATION NUMBER: US/10/081,816  
CURRENT FILING DATE: 2002-02-22  
PRIOR APPLICATION NUMBER: 60/271,319  
PRIOR FILING DATE: 2001-02-23  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 14  
LENGTH: 409  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-10-081-816-14

Query Match 82.8%; Score 24; DB 9; Length 409;  
Best Local Similarity 71.4%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
| | | | |  
Db 24 VLLVAEF 30

RESULT 5  
US-10-024-623-23  
Sequence 23, Application US/10024623  
Publication No. US20020187524A1  
GENERAL INFORMATION:  
APPLICANT: Curtis, Rory A.J.  
TITLE OF INVENTION: 8099, 46455, 54414, 53763, 67076, 67102, 44181,  
TITLE OF INVENTION: 67084FL, AND 67084 ALT, HUMAN PROTEINS AND METHODS OF  
TITLE OF INVENTION: USE THEREOF  
FILE REFERENCE: MNI-214CP  
CURRENT APPLICATION NUMBER: US/10/024,623  
CURRENT FILING DATE: 2001-12-17  
PRIOR APPLICATION NUMBER: US 60/256,240  
PRIOR FILING DATE: 2000-12-15  
PRIOR APPLICATION NUMBER: US 60/256,588  
PRIOR FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: US 60/258,028  
PRIOR FILING DATE: 2000-12-21  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 23  
LENGTH: 1084  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-024-623-23

Query Match 82.8%; Score 24; DB 9; Length 1084;  
Best Local Similarity 71.4%; Pred. No. 3.9e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
: | | | | |  
Db 1016 LMVVAEF 1022

## RESULT 6

US-10-024-623-26  
; Sequence 26, Application US/10024623  
; Publication No. US20020187524A1  
; GENERAL INFORMATION:  
; APPLICANT: Curtis, Rory A.J.  
; TITLE OF INVENTION: 8099, 46455, 54414, 53763, 67076, 67102, 44181,  
; 67084FL, AND 67084 ALT, HUMAN PROTEINS AND METHODS OF  
; TITLE OF INVENTION: USE THEREOF  
; FILE REFERENCE: MNI-214CP  
; CURRENT APPLICATION NUMBER: US/10/024,623  
; CURRENT FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/256,240  
; PRIOR FILING DATE: 2000-12-15  
; PRIOR APPLICATION NUMBER: US 60/256,588  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: US 60/258,028  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 26  
; LENGTH: 1095  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-024-623-26

Query Match 82.8%; Score 24; DB 9; Length 1095;  
Best Local Similarity 71.4%; Pred. No. 4e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
: | | | | |  
Db 1016 LMVVAEF 1022

## RESULT 7

US-10-024-623-36  
; Sequence 36, Application US/10024623  
; Publication No. US20020187524A1  
; GENERAL INFORMATION:  
; APPLICANT: Curtis, Rory A.J.  
; TITLE OF INVENTION: 8099, 46455, 54414, 53763, 67076, 67102, 44181,  
; 67084FL, AND 67084 ALT, HUMAN PROTEINS AND METHODS OF  
; TITLE OF INVENTION: USE THEREOF  
; FILE REFERENCE: MNI-214CP  
; CURRENT APPLICATION NUMBER: US/10/024,623  
; CURRENT FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 60/256,240  
; PRIOR FILING DATE: 2000-12-15  
; PRIOR APPLICATION NUMBER: US 60/256,588  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: US 60/258,028  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 36  
; LENGTH: 1095  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 801..1005  
; OTHER INFORMATION: Xaa = any amino acid  
US-10-024-623-36

Query Match 82.8%; Score 24; DB 9; Length 1095;  
Best Local Similarity 71.4%; Pred. No. 4e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
: | | | | |  
Db 1016 LMVVAEF 1022

## RESULT 8

US-09-158-722-18  
; Sequence 18, Application US/09158722  
; Publication No. US20030013848A1  
; GENERAL INFORMATION:  
; APPLICANT: Lemke Ph.D. et al., Greg E.  
; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/158,722  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/456,647  
; FILING DATE: 02-JUN-1995  
; APPLICATION NUMBER: US 08/237,401  
; FILING DATE: 02-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/884,486  
; FILING DATE: 15-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wetherell Ph.D., John R.  
; REGISTRATION NUMBER: 31,678  
; REFERENCE/DOCKET NUMBER: 07251/007002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 678-5070  
; TELEFAX: (619) 678-5099  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 54 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-158-722-18

Query Match 79.3%; Score 23; DB 9; Length 54;  
Best Local Similarity 57.1%; Pred. No. 31;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
: | | | | |  
Db 9 VMKIADF 15

## RESULT 9

US-09-158-722-46  
; Sequence 46, Application US/09158722  
; Publication No. US20030013848A1  
; GENERAL INFORMATION:  
; APPLICANT: Lemke Ph.D. et al., Greg E.  
; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES

```
;
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/158,722
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/456,647
; FILING DATE: 02-JUN-1995
; APPLICATION NUMBER: US 08/237,401
; FILING DATE: 02-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/884,486
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell Ph.D., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: 07251/007002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-158-722-46

Query Match 79.3%; Score 23; DB 9; Length 54;
Best Local Similarity 57.1%; Pred. No. 31;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VMXVAEF 7
Db 9 VMKIADF 15

RESULT 10
US-09-158-722-47
; Sequence 47, Application US/09158722
; Publication No: US20030013848A1
; GENERAL INFORMATION:
; APPLICANT: Lemke Ph.D. et al., Greg E.
; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/158,722
; FILING DATE:
```

```
;
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/456,647
; FILING DATE: 02-JUN-1995
; APPLICATION NUMBER: US 08/237,401
; FILING DATE: 02-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/884,486
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell Ph.D., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: 07251/007002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-158-722-47

Query Match 79.3%; Score 23; DB 9; Length 54;
Best Local Similarity 57.1%; Pred. No. 31;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VMXVAEF 7
Db 9 VMKIADF 15

RESULT 11
US-09-510-332-40
; Sequence 40, Application US/09510332
; Publication No. US2003002278A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliott
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: T2R, a No. US2003002278A1 Family of Taste Receptors
; FILE REFERENCE: 02307E-0980100S
; CURRENT APPLICATION NUMBER: US/09/510,332
; CURRENT FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 09/393,634
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 40
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human T2R22 (hgr22)
; NAME/KEY: MOD_RES
; LOCATION: (1)..(90)
; OTHER INFORMATION: xaa = any amino acid
; US-09-510-332-40

Query Match 79.3%; Score 23; DB 9; Length 90;
Best Local Similarity 42.9%; Pred. No. 53;
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VMXVAEF 7
Db 11 ILAIAEF 17
```

```
RESULT 12
US-09-393-634-76
; Sequence 76, Application US/09393634
; Patent No. US20020051997A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoop, Mark
; APPLICANT: The Regents of the University of California
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: SF, a No. US20020051997A1el Family of Taste Receptors
; FILE REFERENCE: 02307E-098000US
; CURRENT APPLICATION NUMBER: US/09/393,634
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 76
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human GR22
; NAME/KEY: MOD_RES
; LOCATION: (1)_(90)
; OTHER INFORMATION: xaa - any amino acid
US-09-393-634-76

Query Match          79.3%; Score 23; DB 10; Length 90;
Best Local Similarity 42.9%; Pred. No. 53;
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7
   : : : : :
Db 11 ILAIAEF 17

RESULT 13
US-09-815-242-4899
; Sequence 4899, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4899
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-4899

Query Match          79.3%; Score 23; DB 10; Length 150;
Best Local Similarity 57.1%; Pred. No. 91;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7
   : : : : :
Db 40 LMDIAEF 46

RESULT 14
US-09-738-626-4458
; Sequence 4458, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4458
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4458

Query Match          79.3%; Score 23; DB 9; Length 237;
Best Local Similarity 42.9%; Pred. No. 1.5e+02;
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7
   : : : : :
Db 74 ILSIAEF 80

RESULT 15
US-09-738-626-4505
; Sequence 4505, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
```

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4505  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4505

Query Match 79.3%; Score 23; DB 9; Length 280;  
Best Local Similarity 71.4%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VMXVAEF 7  
I: ||||  
Db 6 VVVVAEF 12

Search completed: April 23, 2003, 12:49:50  
Job time : 15.9655 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:35:24 ; Search time 11.3448 Seconds  
(without alignments)  
18.155 Million cell updates/sec

Title: US-09-723-722A-78  
Perfect score: 29  
Sequence: 1 VMXVAEF 7

Scoring table: BLOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	82.8	620	4	US-08-637-670-40
2	24	82.8	927	4	US-09-134-001C-4831
3	24	82.8	972	3	US-08-335-844A-24
4	23	79.3	54	2	US-08-456-647B-18
5	23	79.3	54	2	US-08-456-647B-46
6	23	79.3	54	2	US-08-456-647B-47
7	23	79.3	54	2	US-08-237-401A-18
8	23	79.3	54	2	US-08-237-401A-46
9	23	79.3	54	2	US-08-237-401A-47
10	23	79.3	299	2	US-08-701-191A-13
11	23	79.3	300	2	US-08-701-191A-31
12	23	79.3	309	2	US-08-701-191A-9
13	23	79.3	310	2	US-08-701-191A-1
14	23	79.3	310	2	US-08-701-191A-6
15	23	79.3	310	2	US-08-701-191A-7
16	23	79.3	310	2	US-08-701-191A-8
17	23	79.3	310	4	US-09-390-326-7
18	23	79.3	313	1	US-08-278-089A-17
19	23	79.3	313	2	US-08-838-957A-16
20	23	79.3	315	2	US-08-701-191A-2
21	23	79.3	318	4	US-08-701-191A-10
22	23	79.3	344	4	US-09-134-001C-5567
23	23	79.3	351	2	US-08-701-191A-3
24	23	79.3	378	1	US-08-070-165F-8
25	23	79.3	378	2	US-08-885-418-8
26	23	79.3	388	1	US-08-070-165F-4
27	23	79.3	388	2	US-08-885-418-4

28	79.3	431	4	US-09-134-001C-4257	Sequence 4257, Ap
29	79.3	612	2	US-08-673-789-11	Sequence 11, Appl
30	79.3	652	1	US-08-471-570-10	Sequence 10, Appl
31	79.3	729	1	US-07-640-029-3	Sequence 3, Appl
32	79.3	729	1	US-08-070-165F-6	Sequence 6, Appl
33	79.3	729	2	US-08-885-418-6	Sequence 5, Appl
34	79.3	731	1	US-07-921-807B-5	Sequence 10, Appl
35	79.3	731	1	US-08-070-165F-10	Sequence 10, Appl
36	79.3	731	1	US-08-441-944A-5	Sequence 5, Appl
37	79.3	731	2	US-08-885-418-10	Sequence 10, Appl
38	79.3	731	4	US-08-439-992A-3	Sequence 3, Appl
39	79.3	733	1	US-07-640-029-4	Sequence 4, Appl
40	79.3	733	1	US-07-921-807B-6	Sequence 6, Appl
41	79.3	733	1	US-08-441-944A-6	Sequence 6, Appl
42	79.3	733	4	US-08-439-992A-4	Sequence 4, Appl
43	79.3	769	1	US-08-471-570-8	Sequence 8, Appl
44	79.3	801	4	US-09-383-630-6	Sequence 6, Appl
45	79.3	802	4	US-09-173-151A-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1  
US-08-637-670-40  
; Sequence 40, Application US/08637670  
; Patent No. 6413521  
; GENERAL INFORMATION:  
; APPLICANT: MCMICHAEL-PHILLIPS et al.  
; TITLE OF INVENTION: Helminth Parasite Antigen with  
; TITLE OF INVENTION: Aminopeptidase-like Activity  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Barbara G. Ernst  
; STREET: 555 13TH STREET, NW Suite 701E  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/637,670  
; APPLICATION NUMBER: US/08/637,670  
; FILING DATE: 26-JUN-1996  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ernst, Barbara G.  
; REGISTRATION NUMBER: 30,377  
; REFERENCE/DOCKET NUMBER: 1811-0232  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-783-6040  
; TELEFAX: 202-783-6031  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 620 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-637-670-40

Query Match 82.8%; Score 24; DB 4; Length 620;  
Best Local Similarity 71.4%; Pred. No. 2.6e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VMXVAEF 7  
I: ||||  
Db 203 VISVAEF 209

RESULT 2  
US-09-134-001C-4831  
; Sequence 4831, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4831  
; LENGTH: 927  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4831

Query Match 82.8%; Score 24; DB 4; Length 927;  
Best Local Similarity 83.3%; Pred. No. 4e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 MXVAEF 7  
Db 126 MSVAEF 131

RESULT 3  
US-08-335-844A-24  
; Sequence 24, Application US/08335844A  
; Patent No. 6066503  
; GENERAL INFORMATION:  
; APPLICANT: GRAHAM, MARGARET  
; APPLICANT: SMITH, TREVOR STANLEY  
; APPLICANT: MUNN, EDWARD ALBERT  
; APPLICANT: KNOX, DAVID PATRICK  
; APPLICANT: OLIVER, JOANNA JANE  
; APPLICANT: NEWTON, SUSAN ELIZABETH  
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING  
; TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF  
; TITLE OF INVENTION: VACCINES AGAINST HELMINTH INFECTIONS  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz  
; STREET: Suite 701-E, 555 Thirteenth St., N.W  
; CITY: Washington  
; STATE: D. C.  
; COUNTRY: U.S.A.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/335.844A  
; FILING DATE: 09-JAN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB PCT/GB93/00943  
; FILING DATE: 06-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9209936  
; FILING DATE: 08-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WALKER, Barbara W.  
; REGISTRATION NUMBER: 35,400  
; REFERENCE/DOCKET NUMBER: 1181-223A  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)783-6040  
; TELEFAX: (202)783-6031  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 972 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-335-844A-24

Query Match 82.8%; Score 24; DB 3; Length 972;  
Best Local Similarity 71.4%; Pred. No. 4.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
Db 542 VISVAEF 548

RESULT 4  
US-08-456-647B-18  
; Sequence 18, Application US/08456647B  
; Patent No. 5811516  
; GENERAL INFORMATION:  
; APPLICANT: Lemke Ph.D. et al., Greg E.  
; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/456,647B  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/237,401  
; FILING DATE: 02-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/884,486  
; FILING DATE: 15-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wetherell Ph.D., John R.  
; REGISTRATION NUMBER: 31,678  
; REFERENCE/DOCKET NUMBER: 07251/007002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 678-5070  
; TELEFAX: (619) 678-5099  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 54 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-456-647B-18

Query Match 79.3%; Score 23; DB 2; Length 54;  
Best Local Similarity 57.1%; Pred. No. 34;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
Db 9 VMKIADF 15

## RESULT 5

US-08-456-647B-46  
; Sequence 46, Application US/08456647B  
; Patent No. 5811516  
; GENERAL INFORMATION:  
; APPLICANT: Lemke Ph.D. et al., Greg E.  
; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/456,647B  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/237,401  
; FILING DATE: 02-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/884,486  
; FILING DATE: 15-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wetherell Ph.D., John R.  
; REGISTRATION NUMBER: 31,678  
; REFERENCE/DOCKET NUMBER: 07251/007002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 678-5070  
; TELEFAX: (619) 678-5099  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 54 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-456-647B-46

Query Match 79.3%; Score 23; DB 2; Length 54;  
Best Local Similarity 57.1%; Pred. No. 34;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
|| :|:|  
DB 9 VMKIADF 15

## RESULT 6

US-08-456-647B-47  
; Sequence 47, Application US/08456647B  
; Patent No. 5811516  
; GENERAL INFORMATION:  
; APPLICANT: Lemke Ph.D. et al., Greg E.  
; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/237,401A  
; FILING DATE: 02-MAY-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/884,486  
; FILING DATE: 15-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile Ph.D., Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07251/007001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 678-5070  
; TELEFAX: (619) 678-5099  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:

; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/456,647B  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/237,401  
; FILING DATE: 02-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/884,486  
; FILING DATE: 15-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wetherell Ph.D., John R.  
; REGISTRATION NUMBER: 31,678  
; REFERENCE/DOCKET NUMBER: 07251/007002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 678-5070  
; TELEFAX: (619) 678-5099  
; INFORMATION FOR SEQ ID NO: 47:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 54 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-456-647B-47

Query Match 79.3%; Score 23; DB 2; Length 54;  
Best Local Similarity 57.1%; Pred. No. 34;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
|| :|:|  
DB 9 VMKIADF 15

## RESULT 7

US-08-237-401A-18  
; Sequence 18, Application US/08237401A  
; Patent No. 5837448  
; GENERAL INFORMATION:  
; APPLICANT: Lemke Ph.D. et al., Greg E.  
; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/237,401A  
; FILING DATE: 02-MAY-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/884,486  
; FILING DATE: 15-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile Ph.D., Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07251/007001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 678-5070  
; TELEFAX: (619) 678-5099  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 54 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-237-401A-18

Query Match 79.3%; Score 23; DB 2; Length 54;  
Best Local Similarity 57.1%; Pred. No. 34;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
DB 9 VMKIADF 15

RESULT 8  
US-08-237-401A-46  
Sequence 46, Application US/08237401A  
Patent No. 5837448  
GENERAL INFORMATION:  
APPLICANT: Lemke Ph.D. et al., Greg E.  
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/237.401A  
FILING DATE: 02-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/884,486  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Halle Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07251/007001  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-237-401A-46

Query Match 79.3%; Score 23; DB 2; Length 54;  
Best Local Similarity 57.1%; Pred. No. 34;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
DB 9 VMKIADF 15

RESULT 9  
US-08-237-401A-47  
Sequence 47, Application US/08237401A  
Patent No. 5837448  
GENERAL INFORMATION:  
APPLICANT: Lemke Ph.D. et al., Greg E.

TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/237.401A  
FILING DATE: 02-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/884,486  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Halle Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07251/007001  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-237-401A-47

Query Match 79.3%; Score 23; DB 2; Length 54;  
Best Local Similarity 57.1%; Pred. No. 34;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
DB 9 VMKIADF 15

RESULT 10  
US-08-701-191A-13  
Sequence 13, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
and Stevan R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/701.191A  
FILING DATE: August 21, 1996

```

; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 227/088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 299 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-701-191A-13

Query Match 79.3%; Score 23; DB 2; Length 299;
Best Local Similarity 57.1%; Pred. No. 2.1e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7
Db 172 VKIADF 178

RESULT 11
US-08-701-191A-31
; Sequence 31, Application US/08701191A
; Patent No. 5942428
; GENERAL INFORMATION:
; APPLICANT: Moosa Mohammadi, Joseph Schlessinger,
; APPLICANT: and Stevan R. Hubbard
; TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN
; TITLE OF INVENTION: OF NON-INSULIN RECEPTOR TYROSINE KINASE
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/701,191A
; FILING DATE: August 21, 1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 227/088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 300 amino acids
; TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-701-191A-31

Query Match 79.3%; Score 23; DB 2; Length 300;
Best Local Similarity 57.1%; Pred. No. 2.1e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7
Db 172 VKIADF 178

RESULT 12
US-08-701-191A-9
; Sequence 9, Application US/08701191A
; Patent No. 5942428
; GENERAL INFORMATION:
; APPLICANT: Moosa Mohammadi, Joseph Schlessinger,
; APPLICANT: and Stevan R. Hubbard
; TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN
; TITLE OF INVENTION: OF NON-INSULIN RECEPTOR TYROSINE KINASE
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/701,191A
; FILING DATE: August 21, 1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 227/088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-701-191A-9

Query Match 79.3%; Score 23; DB 2; Length 309;
Best Local Similarity 57.1%; Pred. No. 2.2e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7
Db 181 VKIADF 187

RESULT 13
US-08-701-191A-1

```

Sequence 1, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
and Stevan R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
SUITE: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08701.191A  
FILING DATE: August 21, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 227/088  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-701-191A-1

Query Match 79.3%; Score 23; DB 2; Length 310;  
Best Local Similarity 57.1%; Pred. No. 2.2e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0;  
Caps 0;

QY 1 VMXVAEF 7  
DB 181 VMKIADF 187

RESULT 14  
US-08-701-191A-6  
Sequence 6, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
and Stevan R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
SUITE: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.

Sequence 7, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
and Stevan R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
SUITE: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08701.191A  
FILING DATE: August 21, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 227/088  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-701-191A-6

Query Match 79.3%; Score 23; DB 2; Length 310;  
Best Local Similarity 57.1%; Pred. No. 2.2e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0;  
Caps 0;

QY 1 VMXVAEF 7  
DB 181 VMKIADF 187

RESULT 15  
US-08-701-191A-7  
Sequence 7, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
and Stevan R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
SUITE: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08701.191A  
FILING DATE: August 21, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.

REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 227/088  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-701-191A-7

Query Match 79.3%; Score 23; DB 2; Length 310;  
Best Local Similarity 57.1%; Pred. No. 2.2e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
|| :||  
Db 181 VMKIADF 187

Search completed: April 23, 2003, 12:46:28  
Job time : 12.3448 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:33:39 ; Search time 21.4828 Seconds  
(without alignments)  
43.419 Million cell updates/sec

Title: US-09-723-722A-78  
Perfect score: 29  
Sequence: 1 VMXVAEF 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	93.1	7	21 AAB07871	A beta-secretase i
2	27	93.1	8	21 AAB07872	A beta-secretase i
3	27	93.1	802	23 ABB93074	Herbicidally activ
4	26	89.7	821	20 AAY34479	Porphyromonas ging
5	26	89.7	869	20 AAY34354	Porphyromonas ging
6	24	82.8	64	23 ABP03038	Human OREX protein
7	24	82.8	75	22 AAU56778	Propionibacterium
8	24	82.8	132	22 AAM41984	Human polypeptide
9	24	82.8	132	22 AAM41985	Human polypeptide
10	24	82.8	132	22 AAM41986	Human polypeptide

11	24	82.8	225	19	AA585820	S. pneumoniae deri
12	24	82.8	236	22	AAU50802	Propionibacterium
13	24	82.8	254	20	AA536930	Amino acid sequenc
14	24	82.8	286	22	ARG18334	Novel human diagno
15	24	82.8	335	22	AAU37753	Streptococcus pneu
16	24	82.8	361	21	AAG20945	Arabisopsis thalia
17	24	82.8	361	21	AAG24458	Arabisopsis thalia
18	24	82.8	361	21	AAG45883	Arabisopsis thalia
19	24	82.8	373	22	AAW40200	Human polypeptide
20	24	82.8	378	21	AAG20944	Arabisopsis thalia
21	24	82.8	378	21	AAG24457	Arabisopsis thalia
22	24	82.8	378	21	AAG45882	Arabisopsis thalia
23	24	82.8	384	21	AA571056	Human membrane tra
24	24	82.8	384	22	AAW40198	Human polypeptide
25	24	82.8	384	22	AAW40199	Human polypeptide
26	24	82.8	414	21	AAG20943	Arabisopsis thalia
27	24	82.8	414	21	AAG24456	Arabisopsis thalia
28	24	82.8	414	21	AAG45881	Arabisopsis thalia
29	24	82.8	584	22	AA596296	Putative p. abyssal
30	24	82.8	817	22	ABB70027	Drosophila melanog
31	24	82.8	927	23	ABP39866	Staphylococcus epi
32	24	82.8	962	15	AA58701	Aminoacylase H11
33	24	82.8	972	14	AA51280	Helminth aminopept
34	23	79.3	36	18	AAW27782	UDP-N-acetylglucos
35	23	79.3	54	19	AAW79151	Receptor protein t
36	23	79.3	54	20	AAW81408	Receptor protein t
37	23	79.3	56	23	AAU78556	Residues 592-647 o
38	23	79.3	68	23	ABP31047	Human ORF20 protei
39	23	79.3	84	22	AAU41413	Propionibacterium
40	23	79.3	90	22	AA587752	Human T2R22 amino
41	23	79.3	143	21	AA10816	Archaeoglobus fulg
42	23	79.3	150	22	AAU33403	Enterococcus faeca
43	23	79.3	162	23	AB55092	Lactococcus lactis
44	23	79.3	189	22	AB552817	Escherichia coli p
45	23	79.3	237	22	AA590704	C glutamicum prote

ALIGNMENTS

RESULT 1  
AAB07871  
ID AAB07871 standard; peptide: 7 AA.

AC AAB07871;  
XX  
DT 14-NOV-2000 (first entry)  
XX  
DE A beta-secretase inhibitor peptide.  
XX  
KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KW inhibitor.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 3 /note= "hydroxyethylene"  
XX  
XX WO200047618-A2.

PD 17-AUG-2000.  
XX  
XX 10-FEB-2000; 2000WO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
PR 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;



QY 1 VMXVAEF 7  
 || :|||  
 Db 629 VMSIAEF 635

## RESULT 4

AA34479  
 ID AAY34479 standard; Protein; 821 AA.

XX AC AAY34479;  
 XX DT 25-AUG-1999 (first entry)  
 XX DE Porphyromonas gingivalis protein PG21.  
 XX KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
 XX OS vaccine; antigenic.  
 XX PN Porphyromonas gingivalis.  
 XX WO9929870-A1.  
 PD 17-JUN-1999.  
 PF 10-DEC-1998; 98WO-AU01023.  
 PR 04-AUG-1998; 98AU-0005028.  
 PR 10-DEC-1997; 97AU-0000839.  
 PR 31-DEC-1997; 97AU-0001182.  
 PR 30-JAN-1998; 98AU-0001546.  
 PR 10-MAR-1998; 98AU-0002264.  
 PR 09-APR-1998; 98AU-0002911.  
 PR 23-APR-1998; 98AU-0003128.  
 PR 05-MAY-1998; 98AU-0003338.  
 PR 22-MAY-1998; 98AU-0003654.  
 PR 29-JUL-1998; 98AU-0004917.  
 XX (CSLC-) CSL LTD.  
 XX PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;  
 PI Ross BC, Rothel LJ, Webb EA;  
 XX WPI; 1999-385613/32.  
 DR N-PSDB; AAX91697.

XX Antigenic Porphyromonas gingivalis peptides for preventing  
 PT gingivitis  
 PS Claim 1; Page 460-461; 588pp; English.  
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic  
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to  
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the  
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial  
 CC activity with a vaccine mechanism of action. The PG polypeptides can be  
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can  
 CC be used to detect Porphyromonas gingivalis in standard hybridisation  
 CC assays. Porphyromonas gingivalis is involved in periodontal disease  
 CC especially gingivitis.  
 XX SQ Sequence 821 AA;

Query Match 89.7%; Score 26; DB 20; Length 821;  
 Best Local Similarity 85.7%; Pred. No. 2.7e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 || :|||  
 Db 565 VMPVAEF 571

## RESULT 5

AA34354

ID AAY34354 standard; Protein; 869 AA.  
 XX AC AAY34354;  
 XX DT 25-AUG-1999 (first entry)  
 XX DE Porphyromonas gingivalis protein PG21.  
 XX KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
 XX OS vaccine; antigenic.  
 XX PN Porphyromonas gingivalis.  
 XX WO9929870-A1.  
 PD 17-JUN-1999.  
 PF 10-DEC-1998; 98WO-AU01023.  
 PR 04-AUG-1998; 98AU-0005028.  
 PR 10-DEC-1997; 97AU-0000839.  
 PR 31-DEC-1997; 97AU-0001182.  
 PR 30-JAN-1998; 98AU-0001546.  
 PR 10-MAR-1998; 98AU-0002264.  
 PR 09-APR-1998; 98AU-0002911.  
 PR 23-APR-1998; 98AU-0003128.  
 PR 05-MAY-1998; 98AU-0003338.  
 PR 22-MAY-1998; 98AU-0003654.  
 PR 29-JUL-1998; 98AU-0004917.  
 XX (CSLC-) CSL LTD.  
 XX PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;  
 PI Ross BC, Rothel LJ, Webb EA;  
 XX WPI; 1999-385613/32.  
 DR N-PSDB; AAX91572.

XX Antigenic Porphyromonas gingivalis peptides for preventing  
 PT gingivitis  
 PS Claim 1; Page 316-318; 588pp; English.  
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic  
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to  
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the  
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial  
 CC activity with a vaccine mechanism of action. The PG polypeptides can be  
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can  
 CC be used to detect Porphyromonas gingivalis in standard hybridisation  
 CC assays. Porphyromonas gingivalis is involved in periodontal disease  
 CC especially gingivitis.  
 XX SQ Sequence 869 AA;

Query Match 89.7%; Score 26; DB 20; Length 869;  
 Best Local Similarity 85.7%; Pred. No. 2.9e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 || :|||  
 Db 613 VMPVAEF 619

## RESULT 6

ABP03038  
 ID ABP03038 standard; Protein; 64 AA.

XX AC ABP03038;  
 XX DT 24-JUN-2002 (first entry)  
 XX DE Human ORFX protein sequence SEQ ID NO:6058.

XX Human; open reading frame; OREF; gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.  
 XX Homo sapiens.  
 XX WO200192523-A2.  
 XX 06-DEC-2001.  
 XX 29-MAY-2001; 2001WO-US10836.  
 XX 30-MAY-2000; 2000US-206132P.  
 PR 29-AUG-2000; 2000US-228716P.  
 XX (CURA-) CURAGEN CORP.  
 PA Shinkets RA, Leach MD;  
 XX WPI; 2002-106308/14.  
 DR N-PSDB; ABN18790.  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders -  
 XX Disclosure; SEQ ID 6058; 1037pp; English.  
 PS  
 XX The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, OREF, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human OREF  
 CC proteins given in ABP00010 to ABP11500. OREF proteins are useful for  
 CC treating or preventing a pathology associated with an OREF-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with OREF-associated disorder. OREF polynucleotide  
 CC sequences can be used in gene therapy. OREF sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative diseases, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. OREF proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 64 AA;

Query Match 82.8%; Score 24; DB 23; Length 64;  
 Best Local Similarity 71.4%; Pred. No. 57;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 || |||  
 DB 53 VMDVAEY 59

RESULT 7  
 AAU56778

ID AAU56778 standard; Protein; 75 AA.  
 XX  
 AC AAU56778;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #17674.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX WPI; 2001-616774/71.  
 DR N-PSDB; AAS59578.  
 XX  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX Example 1; SEQ ID No 17973; 1069pp; English.  
 PS  
 XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 75 AA;

Query Match 82.8%; Score 24; DB 22; Length 75;  
 Best Local Similarity 71.4%; Pred. No. 68;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
 || |||  
 DB 1 VMAVCEF 7

RESULT 8  
 AAU41984

101 AAM41984 standard; Protein; 132 AA.  
102 AAM41984;  
103 DT (first entry)  
104 XX Human polypeptide SEQ ID NO 6915;  
105 DE  
106 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
107 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
108 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
109 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
110 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
111 KW leukaemia.  
112 OS Homo sapiens.  
113 XX WO200153312-A1.  
114 PN  
115 XX 26-JUL-2001.  
116 XX  
117 XX 26-DEC-2000; 2000WO-US34263.  
118 XX  
119 XX 21-JAN-2000; 2000US-0488725.  
120 PR 25-APR-2000; 2000US-0552317.  
121 PR 09-JUL-2000; 2000US-0598042.  
122 PR 19-JUL-2000; 2000US-0620312.  
123 PR 03-AUG-2000; 2000US-0653450.  
124 PR 14-SEP-2000; 2000US-0662191.  
125 PR 19-OCT-2000; 2000US-0693036.  
126 PR 29-NOV-2000; 2000US-0727344.  
127 XX (HYSE-) HYSEQ INC.  
128 PA  
129 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
130 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
131 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
132 XX  
133 DR WPI; 2001-442253/47.  
134 DR N-PSDB; AAI61141.  
135 XX  
136 PT Novel nucleic acids and polypeptides, useful for treating disorders  
137 PT such as central nervous system injuries -  
138 XX  
139 PS Example 2; SEQ ID NO 6915; 10078pp; English.  
140 XX  
141 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
142 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
143 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
144 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
145 CC of the invention may be used to treat diseases of the peripheral nervous  
146 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
147 CC localised neuropathies and central nervous system diseases, such as  
148 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
149 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
150 CC utilisation of the activities such as: Immune system suppression,  
151 CC Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
152 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
153 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
154 CC C.N.S disorders.  
155 CC Note: The sequence data for this patent did not form part of the printed  
156 CC specification.  
157 XX  
158 SQ Sequence 132 AA;  
159  
160 Query Match 82.8%; Score 24; DB 22; Length 132;  
161 Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
162 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
163  
164 QY 1 VMXVAEF 7  
165 Db 64 LMVVAEF 70

166 RESULT 9  
167 AAM41985  
168 ID AAM41985 standard; Protein; 132 AA.  
169 XX  
170 AC AAM41985;  
171 XX  
172 DT 22-OCT-2001 (first entry)  
173 XX  
174 DE Human polypeptide SEQ ID NO 6916.  
175 XX  
176 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
177 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
178 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
179 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
180 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
181 KW leukaemia.  
182 XX  
183 OS Homo sapiens.  
184 XX WO200153312-A1.  
185 PN  
186 XX 26-JUL-2001.  
187 XX  
188 XX 26-DEC-2000; 2000WO-US34263.  
189 XX  
190 XX 21-JAN-2000; 2000US-0488725.  
191 PR 25-APR-2000; 2000US-0552317.  
192 PR 09-JUL-2000; 2000US-0598042.  
193 PR 19-JUL-2000; 2000US-0620312.  
194 PR 03-AUG-2000; 2000US-0653450.  
195 PR 14-SEP-2000; 2000US-0662191.  
196 PR 19-OCT-2000; 2000US-0693036.  
197 PR 29-NOV-2000; 2000US-0727344.  
198 XX (HYSE-) HYSEQ INC.  
199 PA  
200 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
201 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
202 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
203 XX  
204 DR WPI; 2001-442253/47.  
205 DR N-PSDB; AAI61141.  
206 XX  
207 PT Novel nucleic acids and polypeptides, useful for treating disorders  
208 PT such as central nervous system injuries -  
209 XX  
210 PS Example 2; SEQ ID NO 6916; 10078pp; English.  
211 XX  
212 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
213 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
214 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
215 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
216 CC of the invention may be used to treat diseases of the peripheral nervous  
217 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
218 CC localised neuropathies and central nervous system diseases, such as  
219 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
220 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
221 CC utilisation of the activities such as: Immune system suppression,  
222 CC Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
223 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
224 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
225 CC C.N.S disorders.  
226 CC Note: The sequence data for this patent did not form part of the printed  
227 CC specification.  
228 XX  
229 SQ Sequence 132 AA;  
230  
231 Query Match 82.8%; Score 24; DB 22; Length 132;  
232 Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
233 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
234  
235 QY 1 VMXVAEF 7  
236 Db 64 LMVVAEF 70



AAU50802;  
 27-FEB-2002 (first entry)  
 Propionibacterium acnes immunogenic protein #11698.  
 SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 dermatological; osteopathic; neuroprotectant.  
 Propionibacterium acnes.  
 WO200181581-A2.  
 01-NOV-2001.  
 20-APR-2001; 2001WO-US12865.  
 21-APR-2000; 2000US-199047P.  
 02-JUN-2000; 2000US-208841P.  
 07-JUL-2000; 2000US-216747P.  
 (CORI-) CORIXA CORP.  
 Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 WPI; 2001-616774/71.  
 N-PSDB; AAS59549.  
 Propionibacterium acnes polypeptides and nucleic acids useful for  
 vaccinating against and diagnosing infections, especially useful for  
 treating acne vulgaris -  
 Example 1; SEQ ID No 11997; 1069pp; English.  
 Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 polypeptides. The proteins and their associated DNA sequences are used in  
 the treatment, prevention and diagnosis of medical conditions caused by  
 P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 P. acnes is also involved in infections of bone, joints and the central  
 nervous system, however it is particularly involved in the inflammatory  
 lesions associated with acne vulgaris. A method for detecting the  
 presence or absence of P. acnes in a patient comprises contacting a  
 sample with a binding agent that binds to the proteins of the invention  
 and determining the amount of bound protein in the sample. The  
 polypeptides may be used as antigens in the production of antibodies  
 specific for P. acnes proteins. These antibodies can be used to  
 downregulate expression and activity of P. acnes polypeptides and  
 therefore treat P. acnes infections. The antibodies may also be used as  
 diagnostic agents for determining P. acnes presence, for example, by  
 enzyme linked immunosorbent assay (ELISA).  
 Note: The sequence data for this patent did not form part of the printed  
 specification, but was obtained in electronic format directly from WIPO  
 at ftp.wipo.int/pub/published\_pct\_sequences.  
 Sequence 236 AA;  
 Query Match 82.8%; Score 24; DB 22; Length 236;  
 Best Local Similarity 71.4%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VMXVAEF 7  
 || :|||  
 Db 169 VMELAEF 175  
 RESULT 13  
 AAU36930  
 ID AAU36930 standard; Protein: 254 AA.  
 XX

AC AAY36930;  
 XX 07-OCT-1999 (first entry)  
 DT Amino acid sequence of a Chlamydia trachomatis protein.  
 DE  
 XX  
 KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.  
 XX Chlamydia trachomatis.  
 OS  
 XX WO9928475-A2.  
 PN  
 XX 10-JUN-1999.  
 PD  
 XX 27-NOV-1998; 98WO-IB01939.  
 PF  
 XX 04-NOV-1998; 98US-0107077.  
 PR  
 XX 28-NOV-1997; 97FR-0015041.  
 PR  
 XX 17-DEC-1997; 97FR-0016034.  
 XX (GEST ) GENSET.  
 PA  
 XX Griffais R;  
 PI  
 XX WPI; 1999-371125/31.  
 DR  
 XX Genome sequence of Chlamydia trachomatis  
 PT  
 XX Disclosure; Page 786; 1755pp; English.  
 PS  
 XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
 CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nonendemic trachoma, such as  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases, such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC perihhepatitis, bartholinitis; pneumopathy in breast feeding infants;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.  
 XX  
 SQ Sequence 254 AA;  
 Query Match 82.8%; Score 24; DB 20; Length 254;  
 Best Local Similarity 71.4%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VMXVAEF 7  
 || :|||  
 Db 6 VMIVADF 12  
 RESULT 14  
 ABG18334  
 ID ABG18334 standard; Protein: 286 AA.  
 XX  
 AC ABG18334;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 XX Novel human diagnostic protein #18325.  
 DE  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200175067-A2.  
 XX

PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS82521.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 PS Claim 20; SEQ ID NO 48693; 103pp; English.  
 XX  
 PS The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 286 AA;  
 Query Match 82.8%; Score 24; DB 22; Length 286;  
 Best Local Similarity 71.4%; Pred. No. 2.9e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VMXVAEF 7  
 II III  
 Db 23 VMQVAQF 29  
 RESULT 15  
 AAU37753  
 ID AAU37753 standard; Protein; 335 AA.  
 XX  
 AC AAU37753;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Streptococcus pneumoniae cellular proliferation protein #182.  
 XX  
 KW Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX

PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207272P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 XX (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 WPI; 2001-611495/70.  
 DR N-PSDB; AAS55612.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 XX Example 3; Seq ID No 13346; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 335 AA;  
 Query Match 82.8%; Score 24; DB 22; Length 335;  
 Best Local Similarity 71.4%; Pred. No. 3.4e+02;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 VMXVAEF 7  
 II III  
 Db 224 VMSAAEF 230

Search completed: April 23, 2003, 12:35:18  
 Job time : 23.4828 secs



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:33:39 ; Search time 42.9655 Seconds  
(without alignments)  
43.419 Million cell updates/sec

Title: US-09-723-722A-72  
Perfect score: 63  
Sequence: 1 KTEEISEVNXVAEF 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	96.8	14	21	AA07888
2	60	95.2	19	22	AAE00612
3	54	85.7	18	22	AAE00609
4	54	85.7	19	22	AAE00611
5	54	85.7	20	21	AAV69714
6	54	85.7	20	21	AAV69716
7	54	85.7	20	23	AAV78509
8	54	85.7	30	18	AAW08360
9	54	85.7	30	20	AAV33754
10	54	85.7	30	21	AA07895

11	54	85.7	30	22	AA047264
12	54	85.7	30	23	AB090004
13	54	85.7	30	23	AAW50898
14	54	85.7	30	23	AB075999
15	54	85.7	30	23	AAE16664
16	54	85.7	30	23	AAU74838
17	54	85.7	32	17	AAW04402
18	54	85.7	32	17	AAW04403
19	54	85.7	32	17	AAW04401
20	54	85.7	32	23	AB089999
21	54	85.7	32	23	AB075994
22	54	85.7	32	23	AAE16659
23	54	85.7	32	23	AAU74833
24	54	85.7	33	18	AAW08359
25	54	85.7	33	20	AAV33753
26	54	85.7	33	21	AAV07892
27	54	85.7	33	22	AA047263
28	54	85.7	33	23	AB090001
29	54	85.7	33	23	AAW50893
30	54	85.7	33	23	AAW50895
31	54	85.7	33	23	AB075996
32	54	85.7	33	23	AAE16661
33	54	85.7	33	23	AAU74835
34	54	85.7	39	21	AAV69718
35	54	85.7	42	18	AAW08350
36	54	85.7	42	20	AAV33752
37	54	85.7	42	22	AA047262
38	54	85.7	58	20	AAW98001
39	54	85.7	115	20	AAW97997
40	54	85.7	115	20	AAW98000
41	54	85.7	261	22	AAE00610
42	54	85.7	506	19	AAW61152
43	54	85.7	506	20	AAV33742
44	54	85.7	506	22	AA047258
45	54	85.7	695	18	AAW19504

ALIGNMENTS

RESULT 1  
AA07888  
ID AA07888 standard; peptide; 14 AA.

AC AA07888;

DT 14-NOV-2000 (first entry)

XX A peptide fragment derived from beta-amyloid precursor protein.

DE Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide; amyloid plaque component; Alzheimer's disease; amyloidogenic disease; inhibitor.

OS Homo sapiens.

PN WO200047618-A2.

PD 17-AUG-2000.

PF 10-FEB-2000; 2000WO-US03819.

PR 10-FEB-1999; 99US-0119571.

PR 15-JUN-1999; 99US-0139172.

XX (ELAN-) ELAN PHARM INC.

PI Anderson JP, Basi G, Doane MT, Frigone N, John V, Power M;

PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

XX WPI; 2000-533011/48.

PT Purified beta-secretase protein used in assays to discover inhibitors

*Obvious as non-aberrant modification*

PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease -  
 XX  
 PS Disclosure; Page 12; 121pp; English.  
 XX  
 CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents a peptide derived from beta-amyloid  
 CC precursor protein  
 XX  
 SQ Sequence 14 AA;  
 Query Match 96.8%; Score 61; DB 21; Length 14;  
 Best Local Similarity 92.9%; Pred. No. 8.3e-05;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 KTEEISEVNVAAEF 14  
 | | | | | | | | | |  
 DB 1 KTEEISEVNVAAEF 14  
 | | | | | | | | | |  
 RESULT 2  
 AAE00612  
 ID AAE00612 standard; peptide; 19 AA.  
 XX  
 AC AAE00612;  
 XX  
 DT 02-JUL-2001 (first entry)  
 XX  
 DE pC8C1 APP beta-secretase cleavage site (D/I) mutant.  
 XX  
 KW Amyloid precursor protein; APP; caspase; beta-secretase; cleavage site;  
 KW cysteine protease; apoptosis; caspase expression cassette; metastasis;  
 KW tumour; cathepsin B; urokinase; proliferation; gene therapy;  
 KW interdomain linker; Alzheimer's disease; mutant; mutain.  
 XX  
 OS Unidentified.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 12 /note- "Wild-type Asp substituted by Ile"  
 FT  
 XX WO200129232-A2.  
 PN  
 XX 26-APR-2001.  
 PD  
 XX 19-OCT-2000; 2000WO-US28941.  
 PF  
 XX 20-OCT-1999; 99US-0160559.  
 PR  
 XX 14-AUG-2000; 2000US-0225564.  
 XX  
 PA (SCIO-) SCIOS INC.  
 XX  
 PI Cordell B, Li Y;  
 XX  
 XX WPI; 2001-290920/30.  
 DR  
 XX Novel fusion polypeptide comprising first and second caspase subunit  
 PT separated by cleavage site not associated in nature with caspase  
 PT subunit, useful for cloning gene encoding enzymes involved in  
 PT proteolytic cleavage -  
 XX  
 PS Disclosure; Fig 29A; 116pp; English.  
 XX  
 XX The present amino acid sequence is a pC8C1 plasmid amyloid precursor  
 CC protein (APP) beta-secretase cleavage site (AAE00611) (D/I) mutant. APP

CC sequence containing this mutation is less susceptible to beta-secretase  
 CC cleavage. APP beta-secretase cleavage site is used to construct an  
 CC artificially engineered chimeric cassette comprising human caspase-3 with  
 CC interdomain linker replaced by Swedish mutant beta-secretase cleavage  
 CC site. This modified caspase-3 plays a pivotal role in Alzheimer's  
 CC disease. Caspases are a family of cysteine proteases, that participate  
 CC in the initiation and execution of apoptosis.  
 CC The present invention relates to a method for functional cloning of genes  
 CC encoding proteins or enzymes involved in proteolytic cleavage. The  
 CC invention is based on the use of caspase expression cassettes comprising  
 CC the coding sequence of a proteolytic cleavage site flanked by sequences  
 CC encoding two caspase subunits. A fusion polypeptide comprising a first  
 CC and a second caspase subunit, separated by a cleavage site not associated  
 CC in nature, is useful for cloning gene encoding enzymes involved in  
 CC proteolytic cleavage. An expression cassette containing fusion  
 CC polypeptide is used to identify a mutant cell line deficient in an  
 CC enzyme of interest and is also useful for diagnosis and suppression of  
 CC proliferation or metastases of a tumour cell characterised by  
 CC overexpression of a polypeptide (e.g. Cathepsin B or urokinase,  
 CC selectively expressed in the tumour cells). DNA encoding fusion  
 CC polypeptide is used in gene therapy.  
 XX  
 SQ Sequence 19 AA;  
 Query Match 95.2%; Score 60; DB 22; Length 19;  
 Best Local Similarity 85.7%; Pred. No. 0.00018;  
 Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 KTEEISEVNVAAEF 14  
 | | | | | | | | | |  
 DB 2 KTEEISEVNVAAEF 15  
 | | | | | | | | | |  
 RESULT 3  
 AAE00609  
 ID AAE00609 standard; peptide; 18 AA.  
 XX  
 AC AAE00609;  
 XX  
 DT 02-JUL-2001 (first entry)  
 XX  
 DE Beta-amyloid precursor protein beta-secretase cleavage site mutant.  
 XX  
 KW Beta-amyloid precursor protein; beta-APP; caspase; beta-secretase;  
 KW cysteine protease; apoptosis; caspase expression cassette; metastasis;  
 KW tumour; cathepsin B; urokinase; proliferation; gene therapy;  
 KW interdomain linker; cleavage site; Alzheimer's disease; mutant; mutain.  
 XX  
 OS Unidentified.  
 XX  
 PN WO200129232-A2.  
 XX  
 XX 26-APR-2001.  
 PD  
 XX 19-OCT-2000; 2000WO-US28941.  
 PF  
 XX 20-OCT-1999; 99US-0160559.  
 PR  
 XX 14-AUG-2000; 2000US-0225564.  
 XX  
 PA (SCIO-) SCIOS INC.  
 XX  
 PI Cordell B, Li Y;  
 XX  
 XX WPI; 2001-290920/30.  
 DR  
 XX Novel fusion polypeptide comprising first and second caspase subunit  
 PT separated by cleavage site not associated in nature with caspase  
 PT subunit, useful for cloning gene encoding enzymes involved in  
 PT proteolytic cleavage -  
 XX  
 PS Example 2; Page 26; 116pp; English.  
 XX  
 XX The present sequence is a beta-secretase cleavage site of beta-amyloid

CC precursor protein (beta-APP) mutant found in certain families of  
 CC autosomal dominant form of Alzheimer's disease. This sequence is used to  
 CC construct a chimeric cassette comprising human caspase-3 with interdomain  
 CC linker replaced by this sequence. This modified caspase-3 plays a pivotal  
 CC role in Alzheimer's disease. Caspases are a family of cysteine proteases,  
 CC that participate in the initiation and execution of apoptosis.  
 CC The present invention relates to a method for functional cloning of genes  
 CC encoding proteins or enzymes involved in proteolytic cleavage. The  
 CC invention is based on the use of caspase expression cassettes comprising  
 CC the coding sequence of a proteolytic cleavage site flanked by sequences  
 CC encoding two caspase subunits. A fusion polypeptide comprising a first  
 CC and a second caspase subunit, separated by a cleavage site not associated  
 CC in nature, is useful for cloning gene encoding enzymes involved in  
 CC proteolytic cleavage. An expression cassette containing fusion  
 CC polypeptide is used to identify a mutant cell line deficient in an  
 CC enzyme of interest and is also useful for diagnosis and suppression of  
 CC proliferation or metastases of a tumour cell characterised by  
 CC overexpression of a polypeptide (e.g. Cathepsin B or urokinase,  
 CC selectively expressed in the tumour cells). DNA encoding fusion  
 CC polypeptide is used in gene therapy.

XX Sequence 18 AA;

Query Match 85.7%; Score 54; DB 22; Length 18;  
 Best Local Similarity 85.7%; Pred. No. 0.0022;

Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNKNVAEF 14  
 ||||| |||  
 Db 1 KTEEISEVNDLDAEF 14

#### RESULT 4

AAE00611  
 ID AAE00611 standard; peptide; 19 AA.

AC AAE00611;

DT 02-JUL-2001 (first entry)

DE Amyloid precursor protein (APP) beta-secretase cleavage site.

XX Amyloid precursor protein; APP; caspase; beta-secretase; cleavage site;  
 KW cysteine protease; apoptosis; caspase expression cassette; metastasis;  
 KW tumour; cathepsin B; urokinase; proliferation; gene therapy;  
 KW interdomain linker; Alzheimer's disease.

XX Unidentified.

XX WO200129232-A2.

XX 26-APR-2001.

PF 19-OCT-2000; 2000WO-US28941.

XX 20-OCT-1999; 99US-0160559.

PR 14-AUG-2000; 2000US-0225564.

XX (SCIO-) SCIOS INC.

XX Cordell B, Li Y;

XX WPI; 2001-290920/30.

XX Novel fusion polypeptide comprising first and second caspase subunit  
 PT separated by cleavage site not associated in nature with caspase  
 PT subunit, useful for cloning gene encoding enzymes involved in  
 PT proteolytic cleavage.

XX Disclosure; Fig 28A; 116pp; English.

XX The present amino acid sequence is a beta-secretase cleavage site of  
 CC an amyloid precursor protein (APP). This sequence is used to construct

CC an artificially engineered chimeric cassette comprising human caspase-3  
 CC with interdomain linker replaced by Swedish mutant beta-secretase  
 CC cleavage site. This modified caspase-3 plays a pivotal role in  
 CC Alzheimer's disease. Caspases are a family of cysteine proteases, that  
 CC participate in the initiation and execution of apoptosis.  
 CC The present invention relates to a method for functional cloning of genes  
 CC encoding proteins or enzymes involved in proteolytic cleavage. The  
 CC invention is based on the use of caspase expression cassettes comprising  
 CC the coding sequence of a proteolytic cleavage site flanked by sequences  
 CC encoding two caspase subunits. A fusion polypeptide comprising a first  
 CC and a second caspase subunit, separated by a cleavage site not associated  
 CC in nature, is useful for cloning gene encoding enzymes involved in  
 CC proteolytic cleavage. An expression cassette containing fusion  
 CC polypeptide is used to identify a mutant cell line deficient in an  
 CC enzyme of interest and is also useful for diagnosis and suppression of  
 CC proliferation or metastases of a tumour cell characterised by  
 CC overexpression of a polypeptide (e.g. Cathepsin B or urokinase,  
 CC selectively expressed in the tumour cells). DNA encoding fusion  
 CC polypeptide is used in gene therapy.

XX Sequence 19 AA;

Query Match 85.7%; Score 54; DB 22; Length 19;  
 Best Local Similarity 85.7%; Pred. No. 0.0023;

Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNKNVAEF 14  
 ||||| |||  
 Db 2 KTEEISEVNDLDAEF 15

#### RESULT 5

AAE69714

ID AAY69714 standard; peptide; 20 AA.

XX AAY69714;

XX 11-APR-2000 (first entry)

DE Beta-APP alpha-secretase substrate [NL]-APP(-10,+10).

XX Nootropic; neuroprotective; beta-amyloid precursor protein; metabolism;  
 KW cleavage site; beta-secretase; neurodegenerative disease;  
 KW Alzheimer's disease.

XX Homo sapiens.

OS Synthetic.

XX WO9964587-A1.

XX 16-DEC-1999.

PF 04-JUN-1999; 99WO-FR01326.

XX 05-JUN-1998; 98PR-0007068.

PR 31-MAR-1999; 98US-0122599.

XX (RHON) RHONE-POULENC RORER SA.

PA (UYFA-) UNIV CURIE PARIS VI P & M.

XX Rholam M, Munoz-Gimenez N, Moutaouakil M, Cohen P, Bertrand P;

XX WPI; 2000-097537/08.

XX Polypeptide with beta-secretase activity, specific for wild-type  
 PT amyloid precursor protein, useful in treating Alzheimer's disease  
 PT Example 3; Page 24; 44pp; French.

XX Peptides AAY69702-Y69718 represent synthetic peptide substrates for a  
 CC novel polypeptide with beta-secretase activity that can cleave  
 CC specifically the natural beta-amyloid precursor protein (BAPP). Normal  
 CC cleavage of the protein occurs between amino acids Met596-Asp597 and

CC Val636-Ile637 (positions 4-5 and 44-45 of AAY69701). The novel  
 CC polypeptide is used to identify agents that interact specifically with  
 CC it. These agents regulate metabolism of APP, particularly they slow down  
 CC or reduce production of beta-amyloid, so can be used to treat  
 CC neurodegenerative diseases, particularly Alzheimer's disease.  
 XX  
 SQ Sequence 20 AA;

Query Match 85.7%; Score 54; DB 21; Length 20;  
 Best Local Similarity 85.7%; Pred. No. 0.0025;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
 ||||| |||

DB 1 KTEEISEVNXDAEF 14  
 ||||| |||

## RESULT 6

AAY69716  
 ID AAY69716 standard; peptide; 20 AA.

XX AC  
 XX AC  
 XX AAY69716;

DT 11-APR-2000 (first entry)

XX Beta-APP alpha-secretase substrate [NM]-APP(-10,+10).

XX Nootropic; neuroprotective; beta-amyloid precursor protein; metabolism;  
 KW cleavage site; beta-secretase; neurodegenerative disease;  
 KW Alzheimer's disease.

XX Homo sapiens.

OS Synthetic.

PN WO9964587-A1.

XX 16-DEC-1999.

PF 04-JUN-1999; 99WO-FR01326.

XX 05-JUN-1998; 98FR-0007068.

PR 31-MAR-1999; 99US-0122599.

XX (RHON ) RHONE-POULENC RORER SA.

PA (UYPA-) UNIV CURIE PARIS VI P & M.

XX Rholam M, Munoz-Gimenez N, Moutaouakil M, Cohen P, Bertrand P;

PI WPI; 2000-097537/08.

DR Polypeptide with beta-secretase activity, specific for wild-type  
 PT amyloid precursor protein, useful in treating Alzheimer's disease

PS Example 3; Page 24; 44pp; French.

XX Peptides AAY69702-Y69718 represent synthetic peptide substrates for a  
 CC novel polypeptide with beta-secretase activity that can cleave  
 CC specifically the natural beta-amyloid precursor protein (bAPP). Normal  
 CC cleavage of the protein occurs between amino acids Met596-Asp597 and  
 CC Val636-Ile637 (positions 4-5 and 44-45 of AAY69701). The novel  
 CC polypeptide is used to identify agents that interact specifically with  
 CC it. These agents regulate metabolism of APP, particularly they slow down  
 CC or reduce production of beta-amyloid, so can be used to treat  
 CC neurodegenerative diseases, particularly Alzheimer's disease.  
 XX

SQ Sequence 20 AA;

Query Match 85.7%; Score 54; DB 21; Length 20;

Best Local Similarity 85.7%; Pred. No. 0.0025;

Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
 ||||| |||

DB 1 KTEEISEVNXDAEF 14

## RESULT 7

AAU78509  
 ID AAU78509 standard; Peptide; 20 AA.

XX AC  
 XX AAU78509;

DT 18-JUN-2002 (first entry)

XX Beta amyloid precursor protein beta secretase cleavage site.

XX Alzheimer's disease; beta amyloid precursor protein; beta secretase;  
 KW BACE; beta-site APP cleaving enzyme; nootropic; neuroprotective;  
 KW beta-site amyloid precursor protein (APP)-cleaving enzyme; APP;  
 KW BACE secretase/shedase; neurodegenerative disorder.

XX Synthetic.

XX Key Location/Qualifiers

FT Cleavage-site 10..11 /note= "Beta secretase cleavage site"

XX WO200210354-A2.

XX 07-FEB-2002.

XX 01-AUG-2001; 2001WO-CA01118.

XX 01-AUG-2000; 2000CA-2313828.

XX (RECU-) INST RECH CLINIQUES MONTREAL.

XX Seidah NG, Chretien M, Cromlish JA;

XX WPI; 2002-280632/32.

XX Modulating activity of beta-site amyloid precursor protein-cleaving  
 PT enzyme secretase/shedase for treatment of neurodegenerative disorder  
 PT characterised by generation of Abeta protein, by preventing cleavage of  
 PT enzyme

PS Disclosure; Page 21; 64pp; English.

XX This invention relates to a novel method for modulating activity of  
 CC beta-site amyloid precursor protein (APP)-cleaving enzyme (BACE)  
 CC secretase/shedase. Cleavage of BACE by this enzyme results in the  
 CC generation of a soluble BACE which enhances the production of the  
 CC amyloidogenic peptide Abeta which has been shown to be involved in the  
 CC aetiology of Alzheimer's disease. Inhibition of BACE secretase can be  
 CC achieved by administration of an antisense nucleotide molecule capable  
 CC of hybridising with BACE mRNA, by using a ribozyme that targets and  
 CC degrades BACE secretase mRNA, with a peptide that can interfere with  
 CC binding of the enzyme with BACE or using an antibody or antagonist that  
 CC can function as an inhibitor of BACE secretase activation. The methods  
 CC of the invention modulate the activity of BACE secretase/shedase by  
 CC preventing cleavage of BACE, which is useful for the treatment of a  
 CC neurodegenerative disorder characterised by the generation of Abeta  
 CC protein, especially Alzheimer's disease. The invention also comprises a  
 CC method for identification of an agent that can alter the ability of BACE  
 CC secretase to associate with and process a known substrate, this method  
 CC can be used for high throughput screening of candidate molecules. The  
 CC invention also comprises a method for determining whether an individual  
 CC is at risk of developing a neurodegenerative disorder characterised  
 CC by the generation of Abeta protein by measuring the levels of BACE  
 CC C terminal cleavage products in a sample or tissue where an increase  
 CC in cleavage products indicates a person at risk. The present sequence  
 CC represents the Beta secretase cleavage site of the beta amyloid  
 CC precursor protein, this sequence was used to assay beta secretase  
 CC activity of different BACE mutants of the invention.

XX Sequence 20 AA;

```

Query Match      85.7%; Score 54; DB 23; Length 20;
Best Local Similarity 85.7%; Pred. No. 0.0025;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14
   ||||| |||
DB 1 KTEEISEVNLDAEF 14

RESULT 8
AAW08360
ID AAW08360 standard; peptide; 30 AA.
XX AC
XX AC AAW08360;
XX DT 05-SEP-1997 (first entry)
XX DE
XX DE Beta-secretase substrate #2.
XX KW Beta-cleavage site; beta amyloid precursor protein; APP; beta-secretase;
XX KW alpha-secretase; proteolytic cleavage; inhibitor; Alzheimer's disease.
XX OS Synthetic.
XX PN W09640885-A2.
XX PD 19-DEC-1996.
XX PF 07-JUN-1996; 96WO-US09985.
XX PR 07-JUN-1995; 95US-0485152.
XX PR 07-JUN-1995; 95US-0480498.
XX PA (ATHE-) ATHENA NEUROSCIENCES INC.
XX PI Anderson JP, Chrysler SMS, Jacobson-croak KL, Keim PS;
XX PI Mcconlogue LC, Sinha S, Tan H;
XX DR WPI; 1997-052304/05.
XX PT Beta-secretase which specifically cleaves beta-amyloid precursor
XX PT protein - useful to screen for inhibitors useful in treatment of
XX PT Alzheimer's disease
XX PS Disclosure; Page 45; 92pp; English.
XX CC
XX CC AAW08359-W08362 represent substrates for the enzyme of the invention.
XX CC The enzyme of the invention is beta-secretase, and specifically cleaves
XX CC beta-amyloid precursor protein (beta-APP). Normal processing of beta-APP
XX CC is thought to occur via cleavage between residues 16 and 17 of the
XX CC beta-amyloid peptide region by an alpha-secretase. Pathogenic processing
XX CC is thought to occur by beta-secretase cleavage of beta-APP.
XX CC Beta-secretase activity can be detected and measured using a method of
XX CC the invention, which detects at least one of the beta-secretase cleavage
XX CC products formed on cleavage. The method can be used to determine whether
XX CC a test substance inhibits proteolytic cleavage, by beta-secretase, of
XX CC beta-APP. Compounds effective to at least partially inhibit
XX CC beta-secretase activity can be used to inhibit cleavage of beta-APP in
XX CC cells or mammalian hosts. Isolation and purification of beta-secretase
XX CC will permit chemical modelling of a critical event in the pathology of
XX CC Alzheimer's disease.
XX SQ Sequence 30 AA;

Query Match      85.7%; Score 54; DB 18; Length 30;
Best Local Similarity 85.7%; Pred. No. 0.0039;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14
   ||||| |||
DB 17 KTEEISEVNLDAEF 30

RESULT 9
AAV33754
ID AAV33754 standard; Protein; 30 AA.
XX AC
XX AC AAV33754;
XX DT 09-NOV-1999 (first entry)
XX DE
XX DE Synthetic oligopeptide 26-4'SW.
XX KW Beta-secretase; beta-amyloid protein precursor; APP; Down's syndrome;
XX KW Alzheimer's disease; measure activity; cleavage site.
XX OS Synthetic.
XX PN US5942400-A.
XX PD 24-AUG-1999.
XX PF 07-JUN-1996; 96US-0659984.
XX PR 07-JUN-1996; 96US-0659984.
XX PR 07-JUN-1995; 95US-0480498.
XX PR 07-JUN-1995; 95US-0485152.
XX PA (ELAN-) ELAN PHARM INC.
XX PI Anderson JP, Jacobson-Croak KL, Sinha S;
XX PI WPI; 1999-517417/43.
XX PT A method for detecting human beta-secretase cleavage of polypeptides
XX PT useful for identifying beta-secretase inhibitors
XX PS Examples; Column 30; 43pp; English.
XX CC
XX CC Sequences AAV33752-Y33756 are synthetic oligopeptides used for measuring
XX CC the activity of beta-secretase (AAV33741). Beta-secretase is capable of
XX CC cleaving beta-amyloid protein precursor (APP) (AAV33742). These
XX CC synthetic peptides contain the cleavage site of APP. Beta-secretase and
XX CC APP are used in a method for detecting human beta-secretase cleavage of
XX CC polypeptides and for identifying beta-secretase inhibitors. Inhibition
XX CC of beta-secretase activity would be useful for chemical modelling of a
XX CC critical event in the pathology of Alzheimer's disease. Inhibitors of
XX CC beta-secretase would be useful for the prevention and treatment of
XX CC Alzheimer's disease and Down's Syndrome.
XX SQ Sequence 30 AA;

Query Match      85.7%; Score 54; DB 20; Length 30;
Best Local Similarity 85.7%; Pred. No. 0.0039;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14
   ||||| |||
DB 17 KTEEISEVNLDAEF 30

RESULT 10
AAB07895
ID AAB07895 standard; Peptide; 30 AA.
XX AC
XX AC AAB07895;
XX DT 14-NOV-2000 (first entry)
XX DE
XX DE Substrate for beta-secretase enzyme.
XX KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;
XX KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;
XX KW inhibitor.

```

```

OS Synthetic.
XX Key Location/Qualifiers
FH Cleavage-site 26..27
FT
XX WO200047618-A2.
XX
XX 17-AUG-2000.
XX PD
XX PF 10-FEB-2000; 2000WO-US03819.
XX PR 10-FEB-1999; 99US-0119571.
XX PR 15-JUN-1999; 99US-0139172.
XX
XX (ELAN-) ELAN PHARM INC.
XX
XX Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;
PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;
XX
XX WPI; 2000-533011/48.
XX
XX Purified beta-secretase protein used in assays to discover inhibitors
PT which can be used for the treatment of amyloidogenic diseases e.g.
PT Alzheimer's disease -
XX
XX Example 4; Page 71; 121pp; English.
XX
XX The specification describes a beta-secretase enzyme. The enzyme cleaves
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This
CC enzyme is therefore implicated in the production of amyloid plaque
CC components which accumulate in the brains of individuals afflicted with
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's
CC disease-like pathology to test if they maintain or improve cognitive
CC ability or reduce the plaque burden. The compounds are used for the
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The
CC present sequence represents a peptide substrate used to test the
CC activity of beta-secretase enzyme.
XX
XX Sequence 30 AA;
SQ
Query Match 85.7%; Score 54; DB 21; Length 30;
Best Local Similarity 85.7%; Pred. No. 0.0039;
Matches 12; Conservative 0; Mismatches . 2; Indels 0; Gaps 0;
OY 1 KTEEISEVNKXAEF 14
DB 17 KTEEISEVNLDAEF 30
RESULT 11
AAB47264
ID AAB47264 standard; Peptide; 30 AA.
XX
XX AAB47264;
AC
XX 18-JUL-2001 (first entry)
XX
XX Peptide 26-4'SW, for used in beta-secretase assay.
XX
XX Beta-secretase; isotype; beta-amyloid precursor protein; APP;
KW beta-amyloid peptide; beta-AP; Alzheimer's disease; Downs syndrome;
KW HCHWA-D; Swedish mutation; maltose binding protein; MBP.
XX
XX Synthetic.
OS
XX US6221645-B1.
PN
XX
XX 24-APR-2001.
XX PD
XX
XX 07-JUN-1996; 96US-0660531.
XX PF
XX 07-JUN-1995; 95US-0480498.
XX PR

```

```

XX (ELAN-) ELAN PHARM INC.
PA
XX Chrysler SMS, Sinha S, Keim PS, Anderson JP, Tan H, McConlogue LC;
PI
XX WPI; 2001-315578/33.
XX
XX Novel antibody that specifically binds native beta-secretase protein,
PT useful for raising anti-idiotypic antibodies and for detecting or
PT diagnosing pathological conditions related to presence of respective
PT antigens -
XX
XX Example; Column 29; 42pp; English.
XX
XX The sequences given in AAB47262-67 represent synthetic peptides
CC containing the cleavage sites derived from wild-type beta-amyloid
CC precursor protein (APP). These peptides were used in assays utilising
CC partially purified beta-secretase to identify beta-secretase inhibitors.
CC Beta-secretase is thought to be responsible for the pathogenic
CC processing of APP to form beta amyloid peptide (beta-AP) in beta-AP
CC related conditions, e.g. Alzheimer's disease, Downs syndrome, HCHWA-D
CC etc. Beta-secretase has a molecular weight of 260-300 kD and will bind
CC to wheat germ agglutinin but not to concanavalin A. Beta-secretase
CC will cleave both the wild type and the Swedish mutation of APP.
XX
XX Sequence 30 AA;
SQ
Query Match 85.7%; Score 54; DB 22; Length 30;
Best Local Similarity 85.7%; Pred. No. 0.0039;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 KTEEISEVNKXAEF 14
DB 17 KTEEISEVNLDAEF 30
RESULT 12
AAB09004
ID ABB09004 standard; peptide; 30 AA.
XX
XX ABB09004;
AC
XX 19-JUN-2002 (first entry)
DT
XX Peptide #2 used to assay secretase activity.
DE
XX Amyloid precursor protein; APP; Alzheimer's disease; nootropic;
KW neuroprotective; cerebroprotective; haemostatic; antiparkinsonian;
KW Down's syndrome; cerebral amyloid angiopathy; dementia.
XX
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Cleavage-site 26..27
FT /note= "Beta secretase cleavage site"
XX
XX WO200202505-A2.
PN
XX 10-JAN-2002.
XX PD
XX 29-JUN-2001; 2001WO-US20852.
XX PF
XX 30-JUN-2000; 2000US-215323P.
XX PR
XX (ELAN-) ELAN PHARM INC.
PA
XX Fang LY, Hom R, John V, Maillaird M;
XX WPI; 2002-171625/22.
XX
XX New substituted amines are e.g. beta-amyloid peptide inhibitors, useful
PT for treating Alzheimer's disease, mild cognitive impairment and Down's
PT syndrome -

```

XX Example D; Page 97; 136pp; English.  
 XX  
 CC This invention relates to substituted amines and their salts, which  
 CC are nontoxic, neuroprotective, cerebroprotective, haemostatic and  
 CC antiparkinsonian in their action. They are used in the preparation of  
 CC a composition useful for treating, preventing or delaying the onset  
 CC of Alzheimer's disease, for treating mild cognitive impairment, Down's  
 CC syndrome, cerebral amyloid angiopathy, dementia associated with  
 CC Parkinson's disease, dementia associated with progressive supranuclear  
 CC palsy, and dementia associated with cortical basal degeneration.  
 CC They are also useful for treating diseases characterised by beta-amyloid  
 CC deposits in brain, and for producing beta-secretase complexes. This  
 CC sequence represents a peptide that incorporate the known cleavage  
 CC site of beta-secretase. This substrate is used to assay  
 CC beta-secretase activity.  
 XX  
 SQ Sequence 30 AA;  
 Query Match 85.7%; Score 54; DB 23; Length 30;  
 Best Local Similarity 85.7%; Pred. No. 0.0039;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 KTEEISEVNXVAEF 14  
 Db 17 KTEEISEVNLDAEF 30  
 RESULT 13  
 AAM50898  
 ID AAM50898 standard; Peptide; 30 AA.  
 XX  
 AC AAM50898;  
 XX  
 DT 07-MAY-2002 (first entry)  
 XX  
 DE Oligopeptide substrate for beta-secretase.  
 XX  
 KW Beta-secretase; substrate; amyloid precursor protein;  
 KW APP; amyloid beta peptide; inhibitor; screening;  
 KW neurodegenerative disorder; cognitive impairment; Down's syndrome;  
 KW Alzheimer's disease; hereditary cerebral haemorrhage;  
 KW cerebral amyloid angiopathy; dementia; therapy.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Cleavage-site 26...27  
 XX  
 PN WO200202512-A2.  
 XX  
 PD 10-JAN-2002.  
 XX  
 PF 29-JUN-2001; 2001WO-US21012.  
 XX  
 PR 30-JUN-2000; 2000US-215323P.  
 PR 22-NOV-2000; 2000US-252736P.  
 PR 15-DEC-2000; 2000US-255956P.  
 PR 13-FEB-2001; 2001US-268497P.  
 PR 29-MAR-2001; 2001US-279779P.  
 PR 04-JUN-2001; 2001US-295589P.  
 XX  
 PA (ELAN-) ELAN PHARM INC.  
 PA (PHAA ) PHARMACIA & UPJOHN CO.  
 XX  
 PI Maillaird M., Hom C, Gailunas A, Jagodzinska B, Fang LY, John V;  
 PI Freskos JN, Pulley SR, Beck JP, Tenbrink RE;  
 XX  
 DR WPI; 2002-171627/22.  
 XX  
 PT New substituted amines are e.g. amyloid beta peptide production  
 PT inhibitors, useful for treating or preventing e.g. Alzheimer's disease,  
 PT mild cognitive impairment and other degenerative diseases -

XX Example D; Page 344; 651pp; English.  
 XX  
 CC The present sequence is that of a peptide that incorporates the  
 CC known cleavage site of beta-secretase on amyloid precursor protein  
 CC (APP). The peptide, which may optionally be tagged with a  
 CC fluorescent or chromogenic moiety, can be used in assays to  
 CC determine the beta-secretase inhibitory activity of novel  
 CC substituted amine compounds of the invention. Comparison of  
 CC substrate cleavage results in the presence of an inhibitor and  
 CC in the presence of a control provides a measure of the compound's  
 CC inhibitory activity. Novel compounds of the invention are  
 CC effective inhibitors of beta-secretase, inhibit beta-secretase  
 CC mediated cleavage of APP, are effective inhibitors of A beta  
 CC production, and/or effectively reduce amyloid beta deposits or  
 CC plaques. They are used in the treatment or prevention of diseases  
 CC characterised by amyloid deposits in the brain, e.g. for treating  
 CC or preventing Alzheimer's disease (AD), for helping prevent or  
 CC delay the onset of AD, for treating mild cognitive impairment  
 CC (MCI), preventing or delaying the onset of AD in those who would  
 CC progress from MCI to AD, for treating Down's syndrome and  
 CC hereditary cerebral haemorrhage with Amyloidosis of the Dutch  
 CC type, for treating cerebral amyloid angiopathy and preventing its  
 CC potential consequences i.e. single and recurrent lobar haemorrhages,  
 CC for treating other degenerative dementias, including dementias  
 CC of mixed vascular and degenerative origin, dementia associated with  
 CC Parkinson's disease, dementia associated with progressive  
 CC supranuclear palsy, and dementia associated with cortical basal  
 CC degeneration, and for treating diffuse Lewy body type of AD (all  
 CC claimed).  
 XX  
 SQ Sequence 30 AA;  
 Query Match 85.7%; Score 54; DB 23; Length 30;  
 Best Local Similarity 85.7%; Pred. No. 0.0039;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 KTEEISEVNXVAEF 14  
 Db 17 KTEEISEVNLDAEF 30  
 RESULT 14  
 ABB07599  
 ID ABB07599 standard; peptide; 30 AA.  
 XX  
 AC ABB07599;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Synthetic oligopeptide substrate having beta-secretase cleavage site.  
 XX  
 KW APP; amine derivative; beta-secretase; amyloid precursor protein; A beta;  
 KW amyloid beta peptide; beta-amyloid; neurotropic; neuroprotective;  
 KW cerebroprotective; haemostatic; antiparkinsonian.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Cleavage-site 26...27  
 FT /note= "beta-secretase cleavage site"  
 XX  
 PN WO200202520-A2.  
 XX  
 PD 10-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-US21000.  
 XX  
 PR 30-JUN-2000; 2000US-215323P.  
 PR 29-JUN-2001; 2001US-0895843.  
 XX  
 PA (ELAN-) ELAN PHARM INC.  
 PA (PHAA ) PHARMACIA & UPJOHN CO.

XX Beck JP, Gailunas A, Hom R, Jagodzinska B, John V, Maillaird M;  
 PI WPI: 2002-188416/24.  
 XX New substituted amine derivatives, useful for treating Alzheimer's  
 PT disease and other degenerative diseases -  
 XX Example D; Page 163; 286pp; English.  
 XX The invention relates to disubstituted amine derivatives of specified  
 CC formulae or their salts. The amine derivatives can inhibit beta-secretase  
 CC activity or inhibit the cleavage of amyloid precursor protein (APP) in a  
 CC reaction mixture. They can also be used to inhibit production of amyloid  
 CC beta peptide (A beta) in a cell or inhibit production of beta-amyloid  
 CC plaque in an animal. The amine derivatives are useful for treating or  
 CC preventing a disease characterized by beta-amyloid deposits in the brain  
 CC e.g. for treating or preventing Alzheimer's disease (AD), for helping  
 CC prevent or delay the onset of AD, for treating mild cognitive impairment  
 CC (MCI), preventing or delaying the onset of AD in those who would progress  
 CC from MCI to AD, for treating Down's syndrome, humans who have hereditary  
 CC cerebral haemorrhage with the Amyloidosis of the Dutch type, cerebral  
 CC amyloid angiopathy and preventing its potential consequences i.e. single  
 CC and recurrent lobar haemorrhages, for treating other degenerative  
 CC dementias, including dementias of mixed vascular and degenerative origin,  
 CC dementia associated with Parkinson's disease, dementia associated with  
 CC progressive supranuclear palsy, dementia associated with cortical basal  
 CC degeneration, diffuse Lewy body type of Alzheimer's disease. Sequences  
 CC AB807598-599 are synthetic APP oligopeptide substrates containing the  
 CC cleavage site of beta-secretase, used to assay beta-secretase activity.  
 XX  
 SQ Sequence 30 AA:  
 Query Match 85.7%; Score 54; DB 23; Length 30;  
 Best Local Similarity 85.7%; Pred. No. 0.0039;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 KTEEISEVNXVAEF 14  
 ||||| ||||  
 Db 17 KTEEISEVNLDAEF 30  
 RESULT 15  
 AAEL16664  
 ID AAEL16664 standard; peptide; 30 AA.  
 XX  
 AC AAEL16664;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Oligopeptide substrate #2, used in the assay of the invention.  
 XX  
 KW Disubstituted amine derivative; Alzheimer's disease; Down's syndrome;  
 KW degenerative disease; beta-amyloid deposit; mild cognitive impairment;  
 KW hereditary cerebral haemorrhage; amyloidosis; degenerative dementia;  
 KW cerebral amyloid angiopathy; Parkinson's disease; supranuclear palsy;  
 KW cortical basal degeneration; AD; MCI; amyloid precursor protein;  
 KW beta-secretase; amyloid beta peptide; A beta peptide.  
 OS Unidentified.  
 XX  
 XX Key Location/Qualifiers  
 FH Cleavage-site 26..27  
 FT  
 XX  
 PN W0200202518-A2.  
 XX  
 PD 10-JAN-2002.  
 XX  
 XX 29-JUN-2001; 2001WO-US20856.  
 XX  
 XX 30-JUN-2000; 2000US-215323P.  
 XX  
 XX (ELAN-) ELAN PHARM INC.  
 PA

PA (PHAA ) PHARMACIA & UPJOHN CO.  
 XX Beck JP, Gailunas A, Hom R, Jagodzinska B, John V, Maillaird M;  
 XX WPI: 2002-122550/16.  
 XX New substituted amine derivatives, useful for treating Alzheimer's  
 PT disease and other degenerative diseases -  
 XX Example D; Page 163; 286pp; English.  
 XX The invention relates to disubstituted amine derivative compounds  
 CC useful for treating Alzheimer's disease and other degenerative diseases.  
 CC Compounds of the invention possess beta secretase inhibitory activity and  
 CC are useful for inhibiting beta-secretase mediated cleavage of amyloid  
 CC precursor protein (APP) and particularly, to inhibit the production of  
 CC amyloid beta (A beta) peptide. They are useful for treating or preventing  
 CC a disease characterised by beta-amyloid deposits in the brain e.g. for  
 CC treating or preventing Alzheimer's disease (AD), for helping to prevent  
 CC or delay the onset of Alzheimer's disease, for treating mild cognitive  
 CC impairment (MCI), Down's syndrome, hereditary cerebral haemorrhage with  
 CC the amyloidosis of the Dutch type, cerebral amyloid angiopathy and  
 CC preventing its potential consequences i.e. single and recurrent lobar  
 CC haemorrhages, for treating other degenerative dementias including  
 CC dementias of mixed vascular and degenerative origin, dementia associated  
 CC with Parkinson's disease, dementia associated with progressive  
 CC supranuclear palsy, dementia associated with cortical basal degeneration  
 CC and diffuse Lewy body type of Alzheimer's disease. The present sequence  
 CC is oligopeptide substrate used in assay of the invention.  
 XX  
 SQ Sequence 30 AA:  
 Query Match 85.7%; Score 54; DB 23; Length 30;  
 Best Local Similarity 85.7%; Pred. No. 0.0039;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 KTEEISEVNXVAEF 14  
 ||||| ||||  
 Db 17 KTEEISEVNLDAEF 30

Search completed: April 23, 2003, 12:35:16  
 Job time : 43.9655 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:33:40 ; Search time 59.8621 seconds  
(without alignments)  
48.188 Million cell updates/sec

Title: US-09-723-722a-72

Perfect score: 63

Sequence: 1 KTEISEVNXVAEF 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	76.2	35	4 Q8WZ99	Q8WZ99 homo sapien
2	48	76.2	79	11 O35463	O35463 cricetus
3	48	76.2	82	4 Q16014	Q16014 homo sapien
4	48	76.2	82	4 Q16019	Q16019 homo sapien
5	48	76.2	82	4 Q16020	Q16020 homo sapien
6	48	76.2	82	4 P78438	P78438 homo sapien
7	48	76.2	607	11 Q99K32	Q99K32 mus musculus
8	48	76.2	695	6 Q95KN7	Q95KN7 macaca fasc
9	48	76.2	695	11 Q60496	Q60496 cavia sp. p
10	48	76.2	695	11 P97487	P97487 mus musculus
11	48	76.2	770	6 Q9TU10	Q9TU10 sus scrofa
12	47	74.6	534	13 Q93296	Q93296 gallus gall
13	47	74.6	569	13 Q9PVL1	Q9PVL1 gallus gall
14	47	74.6	695	13 Q9DGJ8	Q9DGJ8 gallus gall
15	47	74.6	751	13 Q9DGJ7	Q9DGJ7 gallus gall
16	42	66.7	693	13 Q98SG0	Q98SG0 xenopus lae

Q91963 xenopus. ap  
Q8W0R6 sorghum bic  
O57394 narke japon  
Q9XVK7 caenorhabdi  
Q96847 dictyosteli  
Q23896 dictyosteli  
Q8XHY6 clostridium  
Q8st05 dictyosteli  
Q76251 drosophila  
Q9V6K1 drosophila  
Q9VUL6 drosophila  
Q9XJ88 bacterioph  
Q8th08 fusobacteri  
Q95QW4 caenorhabdi  
Q9AZ64 lactococcus  
Q96LQ3 homo sapien  
Q96LQ3 homo sapien  
Q9F91 arabidopsis  
Q9LYUL arabidopsis  
Q92PB9 rhizobium m  
Q9EM12  
Q9VEF2 drosophila  
Q44228 halocynthia  
Q8SX40 drosophila  
Q9msU0 juniperus c  
Q9msT8 juniperus p  
Q9C8H8 arabidopsis  
Q98821 arabidopsis  
Q9F916 arabidopsis  
O17866 caenorhabdi

747 13 Q91963  
658 10 Q8W0R6  
699 13 O57394  
316 5 Q9XVK7  
335 5 Q96847  
335 5 Q23896  
497 16 Q8XHY6  
1031 5 Q8st05  
1354 5 Q76251  
1354 5 Q9V6K1  
1803 5 Q9VUL6  
213 9 Q9XJ88  
351 16 Q8th08  
561 5 Q95QW4  
564 9 Q9AZ64  
568 4 Q96LQ3  
946 10 Q9F91  
587 10 Q9LYUL  
180 16 Q92PB9  
190 12 Q9EM12  
354 5 Q9VEF2  
415 5 Q44228  
439 5 Q8SX40  
470 8 Q9msU0  
502 8 Q9msT8  
512 10 Q9C8H8  
513 10 Q98821  
570 10 Q9F916  
650 5 O17866

## ALIGNMENTS

### RESULT 1

Q8WZ99 PRELIMINARY; PRT; 35 AA.

AC Q8WZ99;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Amyloid protein (Fragment).  
GN APP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;

RP [1]

RP SEQUENCE FROM N.A.

RA Wakutani Y., Ninomiya H., Iwata H., Tanaka S., Urakami K., Adachi Y.,  
RA Wada-Isoe K., Yamagata K., Ohono K., Tsubuki S., Saido T.,  
RA Hashimoto T., Iwatsubo T., Nakashima K.,  
RT "Novel missense mutation (D678N) of amyloid precursor protein gene in  
RT a Japanese pedigree of familial Alzheimer's disease."  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB066441; BAB71958.1;

FT NON\_TER 1 1

FT NON\_TER 35 35

SQ SEQUENCE 35 AA; 4084 MW; 49D7D17289743B71 CRC64;

Query Match 76.2%; Score 48; DB 4; Length 35;

Best Local Similarity 78.6%; Pred. No. 0.061;

Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14

|||||||

Db 7 KTEISEVNXVAEF 20

|||||||

RESULT 2

O35463 PRELIMINARY; PRT; 79 AA.

ID O35463

AC O35463;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05; Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Alzheimer's amyloid beta protein (Fragment).  
 GN BETA APP.  
 OS Crictetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Crictetus.  
 OX NCBI\_TaxID=10029;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sambamurti K., Pinnix I., Gandhi S.;  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF030413; AAB86608.1; -.  
 DR HSSP; P05067; 1BA4.  
 DR InterPro: IPR001255; Beta-APP.  
 DR Pfam: PF03494; Beta-APP; 1.  
 FT NON\_TER 1  
 FT NON\_TER 79  
 SQ SEQUENCE 79 AA; 8538 MW; 37F2C6C3BFF3F597 CRC64;

Query Match 76.2%; Score 48; DB 11; Length 79;  
 Best Local Similarity 78.6%; Pred. No. 0.14;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KTEEISEVNXVAEF 14  
 ||||| |||  
 Db 11 KTEEISEVKMDAEF 24

## RESULT 3

ID Q16014 PRELIMINARY; PRT; 82 AA.  
 AC Q16014;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Beta-amyloid peptide (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-93236601; PubMed-8476439;  
 RA Denman R.B., Rosenzwaig R., Miller D.L.;  
 RT "A system for studying the effect(s) of familial Alzheimer disease mutations on the processing of the beta-amyloid peptide precursor.";  
 RL Biochem. Biophys. Res. Commun. 192:96-103(1993).  
 DR EMBL; S60721; AAB26263.2; -.  
 DR HSSP; P05067; 1BA4.  
 DR InterPro: IPR001255; Beta-APP.  
 DR Pfam: PF03494; Beta-APP; 1.  
 FT NON\_TER 1  
 FT NON\_TER 82  
 SQ SEQUENCE 82 AA; 8972 MW; F534AA5B3EA9230A CRC64;

Query Match 76.2%; Score 48; DB 4; Length 82;  
 Best Local Similarity 78.6%; Pred. No. 0.15;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KTEEISEVNXVAEF 14  
 ||||| |||  
 Db 8 KTEEISEVKMDAEF 21

## RESULT 4

ID Q16019 PRELIMINARY; PRT; 82 AA.  
 AC Q16019;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Beta-amyloid peptide (Fragment).  
 GN BETA APP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-93236601; PubMed-8476439;  
 RA Denman R.B., Rosenzwaig R., Miller D.L.;  
 RT "A system for studying the effect(s) of familial Alzheimer disease mutations on the processing of the beta-amyloid peptide precursor.";  
 RL Biochem. Biophys. Res. Commun. 192:96-103(1993).  
 DR EMBL; S61380; AAB26264.2; -.  
 DR HSSP; P05067; 1BA4.  
 DR InterPro: IPR001255; Beta-APP.  
 DR Pfam: PF03494; Beta-APP; 1.  
 FT NON\_TER 1  
 FT NON\_TER 82  
 SQ SEQUENCE 82 AA; 8938 MW; F534AA50E579230A CRC64;

Query Match 76.2%; Score 48; DB 4; Length 82;  
 Best Local Similarity 78.6%; Pred. No. 0.15;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KTEEISEVNXVAEF 14  
 ||||| |||  
 Db 8 KTEEISEVKMDAEF 21

## RESULT 5

ID Q16020 PRELIMINARY; PRT; 82 AA.  
 AC Q16020;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Beta-amyloid peptide (Fragment).  
 GN BETA APP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-93236601; PubMed-8476439;  
 RA Denman R.B., Rosenzwaig R., Miller D.L.;  
 RT "A system for studying the effect(s) of familial Alzheimer disease mutations on the processing of the beta-amyloid peptide precursor.";  
 RL Biochem. Biophys. Res. Commun. 192:96-103(1993).  
 DR EMBL; S61383; AAB26265.2; -.  
 DR HSSP; P05067; 1BA4.  
 DR InterPro: IPR001255; Beta-APP.  
 DR Pfam: PF03494; Beta-APP; 1.  
 FT NON\_TER 1  
 FT NON\_TER 82  
 SQ SEQUENCE 82 AA; 8882 MW; F534AA5A5ED9230A CRC64;

Query Match 76.2%; Score 48; DB 4; Length 82;  
 Best Local Similarity 78.6%; Pred. No. 0.15;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KTEEISEVNXVAEF 14  
 ||||| |||  
 Db 8 KTEEISEVKMDAEF 21

## RESULT 6

ID P78438 PRELIMINARY; PRT; 82 AA.  
 AC P78438;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)

DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DR Amyloid protein (Beta-amyloid protein) (Fragment).  
 GN APP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8932030; PubMed=2675837;  
 RA Johnstone E.M., Chaney M.O., Moore R.E., Ward K.E., Norris F.H.,  
 RA Little S.P.;  
 RA "Alzheimer's disease amyloid peptide is encoded by two exons and shows  
 RT similarity to soybean trypsin inhibitor."  
 RL Biochem. Biophys. Res. Commun. 163:1248-1255(1989).  
 RN [2]  
 RP SEQUENCE OF 19-48 FROM N.A.  
 RX MEDLINE=87120329; PubMed=2949367;  
 RA Tanzi R.E., Gusella J.F., Watkins P.C., Bruns G.A., George-Hyslop P.,  
 RA Van Keuren M.L., Patterson D., Pagan S., Kurnit D.M., Neve R.L.;  
 RA "Amyloid beta protein gene: cDNA, mRNA distribution, and genetic  
 RT linkage near the Alzheimer locus."  
 RL Science 235:880-884(1987).  
 RN [3]  
 RP SEQUENCE OF 32-63 FROM N.A.  
 RX MEDLINE=93035397; PubMed=1415269;  
 RA Kamino K., Orr H.T., Payami H., Wijsman E.M., Alonso M.E., Pulst S.M.,  
 RA Anderson L., O'dahl S., Nemens E., White J.A.;  
 RA "Linkage and mutational analysis of familial Alzheimer disease  
 RT kindreds for the APP gene region."  
 RL Am. J. Hum. Genet. 51:998-1014(1992).  
 DR EMBL: M29270; AAAS1768.1; -  
 DR EMBL: M29269; AAAS1768.1; JOINED.  
 DR EMBL: M15532; AAAS1564.1; -  
 DR EMBL: A45136; AAB23646.1; -  
 DR HSP: P05067; IBA4.  
 DR InterPro: IPR001255; Beta-APP.  
 DR Pfam: PF03494; Beta-APP; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 82 AA; 8994 MW; 8DA9E42B813A070E CRC64;  
  
 Query Match 76.2%; Score 48; DB 4; Length 82;  
 Best Local Similarity 78.6%; Pred. No. 0.15;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 KTEEISEVNKVAEF 14  
 ||||| ||||  
 DB 7 KTEEISEVKMDAEF 20  
  
 RESULT 7  
 Q99K32 ID Q99K32 PRELIMINARY; PRT; 607 AA.  
 AC Q99K32;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DE Hypothetical 68.4 kDa protein (Fragment).  
 GN APP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC005490; AAH05490.1; -  
 DR HSP: P05067; IAAp.  
 DR MGD: MGI:88059; App.  
 DR InterPro: IPR001868; A4\_APP.  
 DR InterPro: IPR001255; Beta-APP.

DR InterPro: IPR002223; Kunitz\_BPTI.  
 DR Pfam: PF02177; A4\_EXTRA; 1.  
 DR Pfam: PF03494; Beta-APP; 1.  
 DR Pfam: PF00014; Kunitz\_BPTI; 1.  
 DR PRINTS: PR00203; AMYLOIDA4.  
 DR PRINTS: PR00759; BASICPTASE.  
 DR ProDom: PD000222; Kunitz\_BPTI; 1.  
 DR SMART: SM00131; KU; 1.  
 DR PROSITE: PS00319; A4\_EXTRA; 1.  
 DR PROSITE: PS00320; A4\_INTRA; 1.  
 DR PROSITE: PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE: PS0279; BPTI\_KUNITZ\_2; 1.  
 KW Hypothetical protein; Serine protease inhibitor.  
 FT NON\_TER 1  
 SQ SEQUENCE 607 AA; 68391 MW; BF8022214CBA7D172 CRC64;  
  
 Query Match 76.2%; Score 48; DB 11; Length 607;  
 Best Local Similarity 78.6%; Pred. No. 1.1;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 1 KTEEISEVNKVAEF 14  
 ||||| ||||  
 DB 499 KTEEISEVKMDAEF 512  
  
 RESULT 8  
 Q95KN7 ID Q95KN7 PRELIMINARY; PRT; 695 AA.  
 AC Q95KN7;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Amyloid b-protein precursor.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
 OC Cercopitheidae; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=CEREBELLUM;  
 RX MEDLINE=91273117; PubMed=1905108;  
 RA Podlinsky M.B., Tolani D.R., Selkoe D.J.;  
 RA "Homology of the amyloid beta protein precursor in monkey and human  
 RT supports a primate model for beta amyloidosis in Alzheimer's  
 RT disease".  
 RL Am. J. Pathol. 138:1423-1435(1991).  
 DR EMBL: M58727; AAA36829.1; -  
 DR InterPro: IPR001868; A4\_APP.  
 DR InterPro: IPR001255; Beta-APP.  
 DR Pfam: PF02177; A4\_EXTRA; 1.  
 DR Pfam: PF03494; Beta-APP; 1.  
 DR PROSITE: PS00319; A4\_EXTRA; UNKNOWN\_1.  
 DR PROSITE: PS00320; A4\_INTRA; UNKNOWN\_1.  
 FT SIGNAL 17  
 FT CHAIN 597 636 POTENTIAL.  
 SQ SEQUENCE 695 AA; 78663 MW; 4F6EA0139F969D56 CRC64;  
  
 Query Match 76.2%; Score 48; DB 6; Length 695;  
 Best Local Similarity 78.6%; Pred. No. 1.3;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 1 KTEEISEVNKVAEF 14  
 ||||| ||||  
 DB 587 KTEEISEVKMDAEF 600  
  
 RESULT 9  
 Q60496 ID Q60496 PRELIMINARY; PRT; 695 AA.  
 AC Q60496;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Putative amyloid precursor protein.  
 OS Cavia sp.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
 OX NCBI\_TaxID=10143;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=97236426; PubMed=9116031;  
 RA Beck M., Mueller D., Bigl V.;  
 RT "Amyloid precursor protein in Guinea pigs - complete cDNA sequence and  
 alternative splicing."  
 RL Blochim. Biophys. Acta 1351:17-21(1997).  
 DR EMBL: X97631; CAA66230.1; -;  
 DR HSSP: P05067; IBA4.  
 DR InterPro: IPR001868; A4\_APP.  
 DR InterPro: IPR001255; Beta\_APP.  
 DR Pfam: PF02177; A4\_EXTRA; 1.  
 DR Pfam: PF03494; Beta\_APP; 1.  
 DR PRINTS: PR00203; AMYLOIDA4.  
 DR SMART: SM00006; A4\_EXTRA; 1.  
 DR PROSITE: PS00319; A4\_EXTRA; 1.  
 DR PROSITE: PS00320; A4\_INTRA; 1.  
 SQ SEQUENCE 695 AA; 78701 MW; 5196A0C4017F16AB CRC64;  
  
 Query Match 76.2%; Score 48; DB 11; Length 695;  
 Best Local Similarity 78.6%; Pred. No. 1.3;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 1 KTEEISEVNXVAEF 14  
 |||||  
 DB 587 KTEEISEVKMDAEF 600  
  
 RESULT 10  
 P97487 PRELIMINARY; PRT: 695 AA.  
 AC P97487; p97942;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hippocampal amyloid protein.  
 GN APP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SAMP8; TISSUE=HIPPOCAMPUS;  
 RA Flood J.F., Kumar V.B., Sasser T., Word L., Morley J.E.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 581-662 FROM N.A.  
 RC STRAIN=129SV;  
 RA Wragg M.A., Busfield F., Duff K., Korenblat K., Capecci M.,  
 RA Loring J.F., Goate A.M.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U84012; BAB41502.1; -;  
 DR EMBL: U82624; BAB40519.1; -;  
 DR HSSP: P05067; IMWP.  
 DR MGD: MGI:88059; App.  
 DR InterPro: IPR001868; A4\_APP.  
 DR InterPro: IPR001255; Beta\_APP.  
 DR Pfam: PF02177; A4\_EXTRA; 1.  
 DR Pfam: PF03494; Beta\_APP; 1.  
 DR PRINTS: PR00203; AMYLOIDA4.  
 DR SMART: SM00006; A4\_EXTRA; 1.  
 DR PROSITE: PS00319; A4\_EXTRA; 1.  
 DR PROSITE: PS00320; A4\_INTRA; 1.  
 SQ SEQUENCE 695 AA; 78414 MW; 9A5FBE2ED261236E CRC64;

Query Match 76.2%; Score 48; DB 11; Length 695;  
 Best Local Similarity 78.6%; Pred. No. 1.3;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 1 KTEEISEVNXVAEF 14  
 |||||  
 DB 587 KTEEISEVKMDAEF 600  
  
 RESULT 11  
 Q9TUI0 PRELIMINARY; PRT: 770 AA.  
 AC Q9TUI0;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Amyloid precursor protein.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kimura A., Takahashi T.;  
 RT "Amyloid Precursor Protein 770.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB032550; BAA84580.1; -;  
 DR HSSP: P05067; IAAp.  
 DR InterPro: IPR001868; A4\_APP.  
 DR InterPro: IPR001255; Beta\_APP..  
 DR InterPro: IPR002223; Kunitz\_BPTI.  
 DR Pfam: PF02177; A4\_EXTRA; 1.  
 DR Pfam: PF03494; Beta\_APP; 1.  
 DR Pfam: PF00014; Kunitz\_BPTI; 1.  
 DR PRINTS: PR00203; AMYLOIDA4.  
 DR PRINTS: PR00759; BASICPTASE.  
 DR ProDom: PD000222; Kunitz\_BPTI; 1.  
 DR SMART: SM00006; A4\_EXTRA; 1.  
 DR SMART: SM00131; KU; 1.  
 DR PROSITE: PS00319; A4\_EXTRA; 1.  
 DR PROSITE: PS00320; A4\_INTRA; 1.  
 DR PROSITE: PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE: PS00279; BPTI\_KUNITZ\_2; 1.  
 KW Serine protease inhibitor.  
 SQ SEQUENCE 770 AA; 86961 MW; 5F7A1DCB2BCC583E CRC64;  
  
 Query Match 76.2%; Score 48; DB 6; Length 770;  
 Best Local Similarity 78.6%; Pred. No. 1.4;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 1 KTEEISEVNXVAEF 14  
 |||||  
 DB 662 KTEEISEVKMDAEF 675  
  
 RESULT 12  
 O93296 PRELIMINARY; PRT: 534 AA.  
 AC O93296;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Amyloid protein (Fragment).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98337885; PubMed=9671674;  
 RA Barnes N.Y., Li L., Yoshikawa K., Schwartz L.M., Oppenheim R.W.,  
 RA Milligan C.E.;

RT "Increased production of amyloid precursor protein provides a  
 RL substrate for caspase-3 in dying motoneurons.";  
 J. Neurosci. 18:5869-5880(1998).

DR EMBL; AF042098; AAC25052.1; -  
 DR HSSP; P05067; 1BA4  
 DR InterPro; IPR001868; A4\_APP.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 534 AA; 60597 MW; FB53ECC2E66D4C92 CRC64;

Query Match 74.6%; Score 47; DB 13; Length 534;  
 Best Local Similarity 71.4%; Pred. No. 1.5;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
 ||||:||||  
 Db 426 KTEEVSEVKMDAEF 439

## RESULT 13

Q9PVL1 PRELIMINARY; PRT; 569 AA.  
 AC Q9PVL1;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Amyloid protein (Fragment).  
 GN APP.

OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.

QY 1 KTEEISEVNXVAEF 14  
 ||||:||||  
 RN NCBI\_TaxID=9031;

RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;

RA Coulson E.J., Paliga K., Beyreuther K., Masters C.L.;  
 RT "What the evolution of the amyloid protein precursor superegene family  
 tells us about its function.";  
 RL Neurochem. Int. 0:0-0(2000).  
 DR EMBL; AF030341; AAF12698.1; -  
 DR HSSP; P05067; 1BA4.

DR InterPro; IPR001868; A4\_APP.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR SMART; SM00006; A4\_EXTRA; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 569 AA; 64753 MW; 0AB8BB851863A19D CRC64;

Query Match 74.6%; Score 47; DB 13; Length 569;  
 Best Local Similarity 71.4%; Pred. No. 1.6;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
 ||||:||||  
 Db 462 KTEEVSEVKMDAEF 475

## RESULT 14

Q9DGJ8 PRELIMINARY; PRT; 695 AA.  
 AC Q9DGJ8;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Beta-amyloid precursor protein 695 isoform.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 QY 1 KTEEISEVNXVAEF 14  
 ||||:||||  
 RN NCBI\_TaxID=9031;

RP SEQUENCE FROM N.A.  
 RA Sarasa M., Rodoloso A., Sorribas V.;  
 RT "Cloning of full-length chicken beta-amyloid precursor protein  
 isoforms.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF289218; AAG00593.1; -  
 DR HSSP; P05067; 1BA4  
 DR InterPro; IPR001868; A4\_APP.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR SMART; SM00006; A4\_EXTRA; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 SQ SEQUENCE 695 AA; 78565 MW; F201ED02AEC86D95 CRC64;

Query Match 74.6%; Score 47; DB 13; Length 695;  
 Best Local Similarity 71.4%; Pred. No. 2;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
 ||||:||||  
 Db 587 KTEEVSEVKMDAEF 600

## RESULT 15

Q9DGJ7 PRELIMINARY; PRT; 751 AA.  
 AC Q9DGJ7;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Beta-amyloid precursor protein 751 isoform.  
 OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.

QY 1 KTEEISEVNXVAEF 14  
 ||||:||||  
 RN NCBI\_TaxID=9031;

RP SEQUENCE FROM N.A.  
 RA Sarasa M., Rodoloso A., Sorribas V.;

RT "Cloning of full-length chicken beta-amyloid precursor protein  
 isoforms.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF289219; AAG00594.1; -  
 DR HSSP; P05067; 1BA4.

DR InterPro; IPR001868; A4\_APP.  
 DR InterPro; IPR001255; Beta-APP.  
 DR InterPro; IPR002223; Kunitz\_BPTI.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR Pfam; PF00014; Kunitz\_BPTI; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PROSITE; PS00759; BASICPTASE.  
 DR ProDom; PD000222; Kunitz\_BPTI; 1.  
 DR SMART; SM00006; A4\_EXTRA; 1.  
 DR SMART; SM00131; KU; 1.

DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE; PS00279; BPTI\_KUNITZ\_2; 1.  
 KW Serine protease inhibitor.  
 SQ SEQUENCE 751 AA; 84705 MW; E78E9413A8033D84 CRC64;

Query Match 74.6%; Score 47; DB 13; Length 751;  
Best Local Similarity 71.4%; Pred. NO. 2.2;  
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 KTEEISEVNXVAEF 14  
| | | | | | | | | |  
Db 643 KTEEVEVKMDAEF 656

Search completed: April 23, 2003, 12:38:37  
Job time : 61.8621 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:33:40 ; Search time 9.17241 Seconds  
(without alignments)  
63.306 Million cell updates/sec

Title: US-09-723-722A-72

Perfect score: 63

Sequence: 1 KTEEISEVNXVAEF 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	48	76.2	751	1 A4_SAISC	Q95241 saimiri sci
2	48	76.2	770	1 A4_HUMAN	P05067 homo sapien
3	48	76.2	770	1 A4_MOUSE	P12023 mus musculus
4	48	76.2	770	1 A4_RAT	P08592 rattus norv
5	39	61.9	455	1 YN06_YEAST	P40161 saccharomyc
6	37	58.7	309	1 F5PL_FLABI	P52838 flavaria bi
7	37	58.7	351	1 HRC4_FUSNN	Q87808 fusobacteri
8	37	58.7	1183	1 CNA_STA00	Q53654 staphylococ
9	35	55.6	351	1 FLIC_SERMA	P13713 serratia ma
10	35	55.6	456	1 PUR8_ECOLI	P25739 escherichia
11	35	55.6	473	1 MDW2_XENLA	P56273 xenopus lae
12	35	55.6	505	1 MALQ_STRPN	P29851 streptococc
13	35	55.6	802	1 PAC_ARTVI	P31956 arthrobacte
14	35	55.6	802	1 PAC_BACME	Q60136 bacillus me
15	34	54.0	207	1 RCSA_KLEAE	P05338 klebsiella
16	34	54.0	307	1 YK55_YEAST	P36155 saccharomyc
17	34	54.0	366	1 QUEA_CAUCR	Q98722 caulobacter
18	34	54.0	413	1 ZABB_RABIT	Q00006 o serine/ch
19	34	54.0	426	1 ZABB_PIG	Q29090 s serine/ch
20	34	54.0	427	1 C1SY_ECOLI	P00891 escherichia
21	34	54.0	427	1 C1SY_SALTY	Q68883 salmoneella
22	34	54.0	443	1 ZABB_HUMAN	Q00005 homo sapien
23	34	54.0	443	1 ZABB_PIG	P54614 sus scrofa
24	34	54.0	443	1 ZABB_RAT	P36877 r serine/ch
25	34	54.0	447	1 ZABA_HUMAN	Q00007 h serine/ch
26	34	54.0	447	1 ZABA_RAT	P36876 r serine/ch
27	34	54.0	453	1 ZABD_RAT	P56932 r serine/ch
28	34	54.0	499	1 ZABD_DROME	P36872 drosophila
29	34	54.0	508	1 CC37_CANAL	Q81e6 candida alb
30	34	54.0	520	1 Y870_METUA	P47988 methanococc
31	34	54.0	759	1 TEAL_YEAST	Q58280 saccharomyc
32	34	54.0	967	1 CAP2_MAIZE	P51059 zea mays (m
33	34	54.0	1064	1 YY08_METUA	Q60307 methanococc

34 34 54.0 1612 1 RRPO\_ORSVL  
35 34 54.0 1613 1 VIT2\_CAEEL  
36 34 54.0 1616 1 VIT1\_CAEEL  
37 33 52.4 169 1 CHEL\_ANTPO  
38 33 52.4 234 1 BIOD\_BACSH  
39 33 52.4 263 1 DABP\_LISIN  
40 33 52.4 263 1 DABP\_LISMO  
41 33 52.4 401 1 ODO2\_RICPR  
42 33 52.4 423 1 C1SY\_ACIAN  
43 33 52.4 452 1 YDJE\_ECOLI  
44 33 52.4 663 1 UVRB\_FUSNN  
45 33 52.4 901 1 PIP\_LACLA

#### ALIGNMENTS

RESULT 1

A4\_SAISC  
ID A4\_SAISC STANDARD; PRT: 751 AA.  
AC Q95241;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alzheimer's disease amyloid A4 protein precursor [Contains: Beta-  
DE amyloid protein (Beta-APP) (A-beta)].  
GN APP.  
OS Saimiri sciureus (Common squirrel monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.  
OX NCBI\_TaxID=9521;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver, and Kidney; PubMed=8532114;  
RX MEDLINE=96108492; PubMed=8532114;  
RA Levy E., Amorim A., Frangione B., Walker L.C.;  
RT "Beta-amyloid precursor protein gene in squirrel monkeys with  
RT cerebral amyloid angiopathy."  
RL Neurobiol Aging 16:805-808(1995).  
CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO  
CC -!- INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN  
CC G(O).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION  
CC WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC  
CC RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE  
CC NPXY MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF  
CC PHOSPHORYLATION (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; S81024; AAD14347.1; -.  
CC HSSP; P05067; IAPP.  
CC InterPro; IPR001255; Beta-APP.  
CC InterPro; IPR001223; Kunitz\_BPTI.  
CC Pfam; PF00014; Kunitz\_BPTI; 1.  
CC Pfam; PF02177; A4\_EXTRA; 1.  
CC Pfam; PF03494; Beta-APP; 1.  
CC PRINTS; PR00203; AMYLOIDA4.  
CC PRINTS; PR00759; BASICPTASE.  
CC ProDom; PD000222; Kunitz\_BPTI; 1.  
CC SMART; SM00006; A4\_EXTRA; 1.  
CC SMART; SM00131; KU; 1.  
CC PROSITE; PS00319; A4\_EXTRA; 1.  
CC

P89659 odontogloss  
P05690 caenorhabdi  
P55155 caenorhabdi  
P05686 antheraea p  
P22818 bacillus sp  
Q92aa1 listeria in  
Q8y26 listeria mo  
Q92dy4 rickettsia  
P29002 acinetobact  
P38055 escherichia  
Q8rg2 fusobacteri  
P49022 lactococcus

DR PROSITE; PS00320; A4\_INTRA; 1.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE; PS00279; BPTI\_KUNITZ\_2; 1.  
 KW Glycoprotein; Amyloid; Neurone; Transmembrane; Alternative splicing;  
 KW Signal; Serine protease inhibitor.  
 FT SIGNAL 1  
 FT CHAIN 17 BY SIMILARITY.  
 FT CHAIN 18 751 A4 PROTEIN.  
 FT CHAIN 653 695 BETA-AMYLOID PROTEIN (POTENTIAL).  
 FT DOMAIN 18 680 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 681 704 POTENTIAL.  
 FT DOMAIN 705 751 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 287 345 BPTI/KUNITZ INHIBITOR.  
 FT SITE 740 743 CLATHRIN-BINDING (BY SIMILARITY).  
 FT ACT\_SITE 301 302 REACTIVE BOND.  
 FT ACT\_SITE 291 341 BY SIMILARITY.  
 FT DISULFID 300 324 BY SIMILARITY.  
 FT DISULFID 316 337 BY SIMILARITY.  
 FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (PROBABLE).  
 SQ SEQUENCE 751 AA; 84893 MW; 6C3E431089569049 CRC64;  
 Query Match 76.28; Score 48; DB 1; Length 751;  
 Best Local Similarity 78.66; Pred. NO. 0.38;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 KTEIEISEVNXVAEF 14  
 |||||  
 DB 643 KTEIEISEVKWDAEF 656  
 RESULT 2  
 A4\_HUMAN  
 ID A4\_HUMAN STANDARD; PRT; 770 AA.  
 AC P05067; P09000; Q16011;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alzheimer's disease amyloid A4 protein precursor (Protease nexin-II)  
 DE (PN-II) (APPI) [Contains: Beta-amyloid protein (Beta-APP) (A-beta)].  
 GN APP OR A4 OR CVAP OR ADL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=87144572; PubMed=2881207;  
 RA Kang J., Lemaire H.-G., Unterbeck A., Salbaum J.M., Masters C.L.,  
 RA Grzeschik K.-H., Multhaup G., Beyreuther K., Mueller-Hill B.;  
 RT "The precursor of Alzheimer's disease amyloid A4 protein resembles a  
 RT cell-surface receptor.";  
 RL Nature 325:733-736(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88122639; PubMed=2893289;  
 RA Ponte P., Gonzalez-Dewhitt P., Schilling J., Miller J., Hsu D.,  
 RA Greenberg B., Davis K., Wallace W., Lieberburg I., Fuller F.,  
 RA Cordell B.;  
 RT "A new A4 amyloid mRNA contains a domain homologous to serine  
 RT proteinase inhibitors.";  
 RL Nature 331:525-527(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89128427; PubMed=2783775;  
 RA Lemaire H.G., Salbaum J.M., Multhaup G., Kang J., Bayney R.M.,  
 RA Unterbeck A., Beyreuther K., Mueller-Hill B.;  
 RT "The PreA4(695) precursor protein of Alzheimer's disease A4 amyloid  
 RT is encoded by 16 exons.";  
 RL Nucleic Acids Res. 17:517-522(1989).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97263807; PubMed=9108164;

RA Hattori M., Tsukahara F., Furuhashi Y., Tanahashi H., Hirose M.,  
 RA Saito M., Tsukuni S., Sakaki Y.;  
 RT "A novel method for making nested deletions and its application for  
 RT sequencing of a 300 kb region of human APP locus.";  
 RL Nucleic Acids Res. 25:1802-1808(1997).  
 RN [5]  
 RP SEQUENCE OF 286-345 AND 365-366 FROM N.A.  
 RX MEDLINE=88122640; PubMed=2893290;  
 RA Tanzi R.E., McClatchey A.I., Lamperti E.D., Villa-Komaroff L.,  
 RA Gusella J.F., Neve R.L.;  
 RT "Protease inhibitor domain encoded by an amyloid protein precursor  
 RT mRNA associated with Alzheimer's disease.";  
 RL Nature 331:528-530(1988).  
 RN [6]  
 RP SEQUENCE OF 287-367 FROM N.A.  
 RX MEDLINE=88122641; PubMed=2893291;  
 RA Kitaguchi N., Takahashi Y., Tokushima Y., Shiojiri S., Ito H.;  
 RT "Novel precursor of Alzheimer's disease amyloid protein shows  
 RT protease inhibitory activity.";  
 RL Nature 331:530-532(1988).  
 RN [7]  
 RP SEQUENCE OF 284-289 AND 365-770 FROM N.A.  
 RX MEDLINE=87231971; PubMed=3035574;  
 RA Robakis N.K., Ramakrishna N., Wolfe G., Wisniewski H.M.;  
 RT "Molecular cloning and characterization of a cDNA encoding the  
 RT cerebrovascular and the neuritic plaque amyloid peptides.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4190-4194(1987).  
 RN [8]  
 RP SEQUENCE OF 507-770 FROM N.A.  
 RX MEDLINE=88124954; PubMed=2893379;  
 RA Zain S.B., Salim M., Chou W.G., Sajdel-Sulkowska E.M., Majocha R.E.,  
 RA Marotta C.A.;  
 RT "Molecular cloning of amyloid cDNA derived from mRNA of the Alzheimer  
 RT disease brain: coding and noncoding regions of the fetal precursor  
 RT mRNA are expressed in the cortex.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:929-933(1988).  
 RN [9]  
 RP SEQUENCE OF 672-681.  
 RX MEDLINE=88035004; PubMed=3312495;  
 RA Partridge W.M., Vinters H.V., Yang J., Eisenberg J., Choi T.B.,  
 RA Tourtellotte W.W., Huebner V., Shively J.E.;  
 RT "Amyloid angiopathy of Alzheimer's disease: amino acid composition  
 RT and partial sequence of a 4,200-dalton peptide isolated from cortical  
 RT microvessels.";  
 RL J. Neurochem. 49:1394-1401(1987).  
 RN [10]  
 RP SEQUENCE OF 739-770 FROM N.A.  
 RX MEDLINE=90236318; PubMed=2110105;  
 RA Yoshikai S.-I., Sasaki H., Doh-Ura K., Furuya H., Sakaki Y.;  
 RT "Genomic organization of the human amyloid beta-protein precursor  
 RT gene.";  
 RL Gene 87:257-263(1990).  
 RN [11]  
 RP SEQUENCE OF 1-10 FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=89016647; PubMed=3140222;  
 RA Schon E.A., Mita S., Sadlock J., Herbert J.;  
 RT "A cDNA specifying the human amyloid beta precursor protein (ABPP)  
 RT encodes a 95-kDa polypeptide.";  
 RL Nucleic Acids Res. 16:9351-9351(1988).  
 RN [12]  
 RP SEQUENCE OF 18-50.  
 RX MEDLINE=87250462; PubMed=3597385;  
 RA van Nostrand W.E., Cunningham D.D.;  
 RT "Purification of protease nexin II from human fibroblasts.";  
 RL J. Biol. Chem. 262:8508-8514(1987).  
 RN [13]  
 RP IDENTITY OF APP WITH NEXIN-II.  
 RX MEDLINE=89384866; PubMed=2506449;  
 RA Oltersdorf T., Fritz L.C., Schenk D.B., Lieberburg I.,  
 RA Johnson-Wood K.L., Beattie E.C., Ward P.J., Blacher R.W., Dovey H.F.,  
 RA Sinha S.;  
 RT "The secreted form of the Alzheimer's amyloid precursor protein with



RT the Kunitz domain is protease nexin-II.";  
 RL Nature 341:144-147(1989).  
 RN [14]  
 RP PROTEASE-SPECIFICITY OF INHIBITOR DOMAIN.  
 RX MEDLINE-90211252; PubMed-1969731;  
 RA Kido H., Fukutomi A., Schilling J., Wang Y., Cordell B., Katunuma N.;  
 RT "Protease-specificity of Kunitz inhibitor domain of Alzheimer's  
 RL disease amyloid protein precursor.";  
 RL Biochem. Biophys. Res. Commun. 167:716-721(1990).  
 RN [15]  
 RP COMPLEX WITH G(O).  
 RX MEDLINE-93188965; PubMed-8446172;  
 RA Nishimoto I., Okamoto T., Matsuura Y., Takahashi S., Okamoto T.,  
 RA Murayama Y., Ogata E.;  
 RT "Alzheimer amyloid protein precursor complexes with brain GTP-binding  
 RL protein G(O).";  
 RL Nature 362:75-79(1993).  
 RN [16]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 28-133.  
 RX MEDLINE-99215582; PubMed-10201399;  
 RA Rossjohn J., Cappai R., Feil S.C., Henry A., McKinstry W.J.,  
 RA Galatis D., Hesse L., Multhaup G., Beyreuther K., Masters C.L.,  
 RA Parker M.W.;  
 RT "Crystal structure of the N-terminal, growth factor-like domain of  
 RL Alzheimer amyloid precursor protein.";  
 RL Nat. Struct. Biol. 6:327-331(1999).  
 RN [17]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 287-344.  
 RX MEDLINE-91104913; PubMed-2125487;  
 RA Hynes T.R., Randal M., Kennedy L.A., Eigenbrot C., Kosiakof A.A.;  
 RT "X-ray crystal structure of the protease inhibitor domain of  
 RL Alzheimer's amyloid beta-protein precursor.";  
 RL Biochemistry 29:10018-10022(1990).  
 RN [18]  
 RP STRUCTURE BY NMR OF 289-344.  
 RX MEDLINE-92031488; PubMed-1718421;  
 RA Heald S.L., Tilton R.F. Jr., Hammond L.S., Lee A., Bayney R.M.,  
 RA Karamark M.E., Ramabhadran T.V., Dreyer R.N., Davis G., Unterbeck A.,  
 RA Tamburini P.P.;  
 RT "Sequential NMR resonance assignment and structure determination of  
 RL the Kunitz-type inhibitor domain of the Alzheimer's beta-amyloid  
 RL precursor protein.";  
 RL Biochemistry 30:10467-10478(1991).  
 RN [19]  
 RP STRUCTURE BY NMR OF 672-699.  
 RX MEDLINE-94281210; PubMed-7516706;  
 RA Talafous J., Marciniowski K.J., Klopman G., Zagorski M.G.;  
 RT "Solution structure of residues 1-28 of the amyloid beta-peptide.";  
 RL Biochemistry 33:7788-7796(1994).  
 RN [20]  
 RP STRUCTURE BY NMR OF 696-706.  
 RX MEDLINE-97128622; PubMed-8973180;  
 RA Kohno T., Kobayashi K., Maeda T., Sato K., Takashima A.;  
 RT "Three-dimensional structures of the amyloid beta peptide (25-35) in  
 RL membrane-mimicking environment.";  
 RL Biochemistry 35:16094-16104(1996).  
 RN [21]  
 RP STRUCTURE BY NMR OF 672-711.  
 RX MEDLINE-98359783; PubMed-9693002;  
 RA Coles M., Bicknell W., Watson A.A., Fairlie D.P., Craik D.J.;  
 RT "Solution structure of amyloid beta-peptide(1-40) in a water-micelle  
 RL environment. Is the membrane-spanning domain where we think it is?";  
 RL Biochemistry 37:11064-11077(1998).  
 RN [22]  
 RP STRUCTURE BY NMR OF 672-699.  
 RX MEDLINE-20400066; PubMed-10940222;  
 RA Poulsen S.A., Watson A.A., Craik D.J.;  
 RT "Solution structures in aqueous SDS micelles of two amyloid beta  
 RL peptides of Abeta(1-28) mutated at the alpha-secretase cleavage  
 RL site.";  
 RL J. Struct. Biol. 130:142-152(2000).  
 RN [23]  
 RP STRUCTURE BY NMR OF 681-706.

RX MEDLINE-20400065; PubMed-10940221;  
 RA Zhang S., Iwata K., Lachenmann M.J., Peng J.W., Li S., Stimson E.R.,  
 RA Lu Y., Felix A.M., Maggio J.E., Lee J.P.;  
 RT "The Alzheimer's peptide a beta adopts a collapsed coil structure in  
 RL water.";  
 RL J. Struct. Biol. 130:130-141(2000).  
 RN [24]  
 RP SIGNAL SEQUENCE CLEAVAGE SITE, AND TOPOLOGY.  
 RX MEDLINE-88296437; PubMed-2900137;  
 RA Dyrks T., Weidemann A., Multhaup G., Salbaum J.M., Lemaire H.-G.,  
 RA Kang J., Mueller-Hill B., Masters C.L., Beyreuther K.;  
 RT "Identification, transmembrane orientation and biogenesis of the  
 RL amyloid A4 precursor of Alzheimer's disease.";  
 RT Query Match 76.2%; Score 48; DB 1; Length 770;  
 RT Best Local Similarity 78.6%; Pred. No. 0.39;  
 RT Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 KTEEISEVNXVAEF 14  
 ID A4\_MOUSE STANDARD; PRT: 770 AA.  
 AC P12023;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alzheimer's disease amyloid A4 protein homolog precursor  
 DE {Amyloidogenic glycoprotein} (AG).  
 DE APP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-Brain;  
 RX MEDLINE-92096458; PubMed-1756177;  
 RA de Strooper B., van Leuven F., van den Berghe H.;  
 RT "The amyloid beta protein precursor or proteinase nexin II from mouse  
 RL is closer related to its human homolog than previously reported.";  
 RL Biochim. Biophys. Acta 1129:141-143(1991).  
 RN [2]  
 RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE-88106489; PubMed-3322280;  
 RA Yamada T., Sasaki H., Furuya H., Miyata T., Goto I., Sasaki Y.;  
 RT "Complementary DNA for the mouse homolog of the human amyloid beta  
 RL protein precursor.";  
 RL Biochem. Biophys. Res. Commun. 149:665-671(1987).  
 RN [3]  
 RP REVISIONS.  
 RA Yamada T.;  
 RL Submitted (MAR-1988) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 289-364 FROM N.A.  
 RC STRAIN=CD-1; TISSUE=Placenta;  
 RX MEDLINE-89345111; PubMed-2569710;  
 RA Fukuchi K., Martin G.M., Deeb S.S.;  
 RT "Sequence of the protease inhibitor domain of the A4 amyloid protein  
 RL precursor of Mus domesticus";  
 RL Nucleic Acids Res. 17:5396-5396(1989).  
 RN [5]  
 RP SEQUENCE OF 1-19 FROM N.A.  
 RX MEDLINE-92209998; PubMed-1555768;  
 RA Izumi R., Yamada T., Yoshikai S.I., Sasaki H., Hattori M.,  
 RA Sakai Y.;  
 RT "Positive and negative regulatory elements for the expression of the  
 RL Alzheimer's disease amyloid precursor-encoding gene in mouse.";  
 RL Gene 112:189-195(1992).



Pfam; PF02177; A4\_EXTRA; 1.  
DR PFam; PF03494; Beta-APP; 1.  
DR PRINTS; PR00203; AMYLOIDA4.  
DR PRINTS; PR00759; BASICPTASE.  
DR ProDom; PD000222; Kunitz\_BPTI; 1.  
DR SMART; SM00006; A4\_EXTRA; 1.  
DR SMART; SM00131; KU; 1.  
DR PROSITE; PS00319; A4\_EXTRA; 1.  
DR PROSITE; PS00320; A4\_INTRA; 1.  
DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
DR PROSITE; PS00279; BPTI\_KUNITZ\_2; 1.  
DR Glycoprotein; Amyloid; Neurone; Transmembrane; Signal;  
KW Alternative splicing; Serine protease inhibitor.  
FT SIGNAL 1 17  
FT CHAIN 18 770  
FT ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN  
FT HOMOLOG.  
FT DOMAIN 18 699  
FT TRANSMEM 700 723  
FT DOMAIN 724 770  
FT DOMAIN 673 715  
FT DOMAIN 287 345  
FT SITE 759 762  
FT SITE 291 341  
FT DISULFID 300 324  
FT DISULFID 316 337  
FT CARBOHYD 542 542  
FT CARBOHYD 571 571  
FT VARSPLIC 289 289  
FT VARSPLIC 290 364  
FT MISSING (IN ISOFORM APP(695)).  
SQ SEQUENCE 770 AA; 86704 MW; C26C9D6BB2D929A7 CRC64;  
  
Query Match 76.2%; Score 48; DB 1; Length 770;  
Best Local Similarity 78.6%; Pred. No. 0.39;  
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 KTEEISEVNVXAEF 14  
|||||||  
DB 662 KTEEISEVNVXAEF 675  
  
RESULT 5  
YNU6\_YEAST STANDARD; PRT; 455 AA.  
ID YNU6\_YEAST  
AC P40161;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Hypothetical 51.6 kDa protein in SSB2-SPX18 intergenic region.  
GN YNL206C OR N1346.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / FY1679;  
RX MEDLINE=95242839; PubMed=7725799;  
RA Jonniaux J.-L., Coster F., Purnelle B., Goffeau A.;  
RT "A 21.7 kb DNA segment on the left arm of yeast chromosome XIV  
RT carries WHI3, CCR2, SPX18, SPX19, an homologue to the heat shock gene  
RT SSB1 and 8 new open reading frames of unknown function.";  
RL Yeast 10:1639-1645(1994).  
CC -!- SIMILARITY: SOME, TO DROSOPHILA AND MAMMALIAN SINGLE-STRAND  
CC RECOGNITION PROTEINS (SSRP).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC

EMBL; X78898; CAA55502.1; -.  
DR EMBL; Z71482; CAA96106.1; -.  
DR SGD; S0005150; YNL206C.  
KW Hypothetical protein.  
SQ SEQUENCE 455 AA; 51621 MW; 79CACA659A7F4F4D CRC64;  
  
Query Match 61.9%; Score 39; DB 1; Length 455;  
Best Local Similarity 63.6%; Pred. No. 10;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 KTEEISEVNVX 11  
||:|||||:  
DB 63 KTDEISEVTNTI 73  
  
RESULT 6  
FSTL\_FLABI STANDARD; PRT; 309 AA.  
ID FSTL\_FLABI  
AC P52838;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Flavonol sulfotransferase-like (EC 2.8.2.-).  
OS Flaveria bidentis.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;  
OC Helenieae; Flaveria.  
OX NCBI\_TaxID=4224;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95232186; PubMed=7716232;  
RA Ananvoranich S., Gullick P., Ibrahim R.K.;  
RT "Flavonol sulfotransferase-like cDNA clone from Flaveria bidentis.";  
RL Plant Physiol. 107:1019-1020(1995).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -!- SIMILARITY: BELONGS TO THE PLANT SULFOTRANSFERASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC  
EMBL; U10277; AAA87399.1; -.  
DR HSSP; P50224; ICJM.  
DR InterPro; IPR000863; Sulfotransferase.  
DR Pfam; PF00685; Sulfotransfer; 1.  
DR ProDom; PD001218; Sulfotransferase; 1.  
KW Transferase.  
FT BINDING 138 153 PAPS-BINING SITE (BY SIMILARITY).  
SQ SEQUENCE 309 AA; 35917 MW; 598A9C9EFBB75303 CRC64;  
  
Query Match 58.7%; Score 37; DB 1; Length 309;  
Best Local Similarity 57.1%; Pred. No. 16;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 KTEEISEVNVXAEF 14  
||:|||||:  
DB 211 KSEPSISNVKRLAEF 224  
  
RESULT 7  
HRCA\_FUSNN STANDARD; PRT; 351 AA.  
ID HRCA\_FUSNN  
AC Q8RH08;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Heat-inducible transcription repressor hrca.  
GN HRCA OR FN0113.





```

RT MDM2 protein from Xenopus laevis."
RL Oncogene 14:1427-1433(1997).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 13-119 IN COMPLEX WITH P53.
RX MEDLINE=97081050; PubMed=8875929;
RA Kussie P.H., Gorina S., Marechal V., Elenbaas B., Moreau J.,
RA Levine A.J., Pavletich N.P.;
RT "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor
transactivation domain."
RL Science 274:948-953(1996).
CC -1- FUNCTION: MAY BIND P53 PROTEIN AND MAY FUNCTION AS AN UBIQUITIN
CC LIGASE E3.
CC -1- COFACTOR: ZINC IS REQUIRED FOR UBIQUITIN LIGASE E3 ACTIVITY (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -1- DEVELOPMENTAL STAGE: EXPRESSION INCREASES FROM OOCYTE STAGE I/II
CC TO REACH ITS MAXIMUM IN OOCYTE STAGE V/VI IN UNFERTILIZED EGGS,
CC AND THEN PROGRESSIVELY DECREASES TO BECOME UNDETECTABLE AT THE
CC GASTRULA STAGE.
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -1- SIMILARITY: CONTAINS 1 RANBP2-TYPE ZINC FINGER.
CC -1- SIMILARITY: BELONGS TO THE MDM2 / MDM4 FAMILY.
CC PDB; LYCO; 19-NOV-97.
DR InterPro; IPR003160; MDM2.
DR InterPro; IPR001876; Znf_RanGDP.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00641; zf-RanBP; 1.
DR Pfam; PF02279; MDM2; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00347; Znf_RB2; 1.
DR PROSITE; PS01358; ZF-RANBP2_1; 1.
DR PROSITE; PS00199; ZF-RANBP2_2; 1.
DR PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW Nuclear protein; Ligase; Ub1 conjugation pathway; Zinc; Zinc-finger;
KW Metal-binding; 3D-structure.
FT DOMAIN 15 104 REGION I.
FT DOMAIN 173 179 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 230 322 REGION II.
FT DOMAIN 216 292 ASP/GLU-RICH (ACIDIC).
FT ZN_FING 290 319 RANBP2-TYPE.
FT ZN_FING 420 461 RING-TYPE.
SQ SEQUENCE 473 AA; 53464 MW; 7DA668DE9B3BEE01 CRC64;

Query Match 55.6%; Score 35; DB 1; Length 473;
Best Local Similarity 61.5%; Pred. No. 56;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAE 13
DB 160 ETEISSVDHPAE 172
:|||||:|

RESULT 12
MALQ_STRPN STANDARD; PRT; 505 AA.
AC P29851;
DT 01-APR-1993 (Rel. 25; Created)
DT 01-APR-1993 (Rel. 25; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE 4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase)
DE (disproportionating enzyme) (D-enzyme).
GN MALQ OR MAM OR SP2107.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83129347; PubMed=6297760;
RA Lacks S.A., Dunn J.J., Greenberg B.;
RT "Identification of base mismatches recognized by the
heteroduplex-DNA-repair system of Streptococcus pneumoniae.";
```

```

Cell 31:327-336(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Neilson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Purkerson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfle E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae".
RL Science 293:498-506(2001).
CC -1- CATALYTIC ACTIVITY: Transfers a segment of a (1,4)-alpha-D-glucan
CC to a new 4-position in an acceptor, which may be glucose or (1,4)-
CC alpha-D-glucan.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE DISPROPORTIONATING ENZYME FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J01796; AAA26923.1; -.
CC DR EMBL; AE007499; AAK76166.1; -.
CC DR HSSP; O87172; ICWY.
CC DR TIGR; SP2107;
CC InterPro; IPR003385; 4A_glucanotrans.
CC Pfam; PF02446; 4A_glucanotrans; 1.
CC DR TIGR; TIGR00217; malQ; 1.
KW Transferase; Glycosyltransferase; Carbohydrate metabolism;
KW Complete proteome.
SQ SEQUENCE 505 AA; 58076 MW; D4529A000A6D01A5 CRC64;

Query Match 55.6%; Score 35; DB 1; Length 505;
Best Local Similarity 54.5%; Pred. No. 60;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 EEISEVNXVAE 13
DB 332 EEELGNLIIAE 342
||:|:|:|

RESULT 13
PAC_ARTVI STANDARD; PRT; 802 AA.
AC P31956;
DT 01-JUL-1993 (Rel. 26; Created)
DT 01-JUL-1993 (Rel. 26; Last sequence update)
DT 15-JUN-2002 (Rel. 41; Last annotation update)
DE Penicillin G acylase precursor (EC 3.5.1.11) (Penicillin G amidase)
DE (Penicillin G amidohydrolase).
GN PAC OR PA.
OS Arthrobacter viscosus.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Micrococccaceae; Micrococccaceae; Arthrobacter.
OX NCBI_TaxID=1673;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94259306; PubMed=8200542;
RA Konstantinovic M., Marjanovic N., Ljubijankic G., Glisin V.;
RT "The penicillin amidase of Arthrobacter viscosus (ATCC 15294).";
RL Gene 143:79-83(1994).
CC -1- CATALYTIC ACTIVITY: Penicillin + H(2)O -> a fatty acid anion + 6-
CC aminopenicillanate.
```

CC -1- SUBUNIT: HETERODIMER OF AN ALPHA CHAIN AND A BETA CHAIN  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Extracellular (Potential).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S45.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L04471; AAA22077.1; -  
 CC HSSP; P06875; 1AJQ.  
 CC InterPro: IPR002692; Penicill\_amidase.  
 CC Pfam; PF01804; Penicill\_amidase; 1.  
 CC Hydrolase; Antibiotic resistance; Zymogen; Signal.  
 CC SIGNAL 1 24 POTENTIAL.  
 CC CHAIN 25 802 PENICILLIN G ACYLASE ZYMOGEN.  
 CC CHAIN 25 234 PENICILLIN G ACYLASE ALPHA SUBUNIT.  
 CC PROPEP 235 265 SPACER PEPTIDE.  
 CC CHAIN 266 802 PENICILLIN G ACYLASE BETA SUBUNIT.  
 CC ACT\_SITE 266 266 BY SIMILARITY.  
 CC SEQUENCE 802 AA; 92113 MW; 9863E58C526C85D7 CRC64;  
 CC  
 CC Query Match 55.6%; Score 35; DB 1; Length 802;  
 CC Best Local Similarity 50.0%; Pred. No. 96;  
 CC Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 CC  
 CC QY 3 EEISEVNXVAEF 14  
 CC |:|:|:| | | |  
 CC Db 548 EDINEINTASF 559  
 CC  
 CC RESULT 14  
 CC PAC\_BACME STANDARD; PRT; 802 AA.  
 CC ID PAC\_BACME Q9S463;  
 CC AC Q60136; Q9S463;  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Penicillin G acylase precursor (EC 3.5.1.11) (Penicillin G amidase)  
 CC DE (Penicillin G amidohydrolase).  
 CC GN PAC OR PGA.  
 CC OS Bacillus megaterium.  
 CC OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 CC OX NCBI\_TaxID=1404;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=ATCC 14945;  
 CC RA Kang J.H., Kim S.J., Park Y.C., Hwang Y., Yoo O.J., Kim Y.C.;  
 CC RT "Nucleotide sequence of the penicillin G acylase gene from  
 CC RT Bacillus megaterium and characteristics of the enzyme.";  
 CC RL Misalunmuhag Holji 32:215-221(1994).  
 CC RN [2]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=ATCC 14945;  
 CC RX MEDLINE=95180705; PubMed=7875576;  
 CC RA Martin L.M., Prieto A.M., Cortes E., Garcia J.L.;  
 CC RT "Cloning and sequencing of the pac gene encoding the penicillin G  
 CC RT acylase of Bacillus megaterium ATCC 14945.";  
 CC RL FEMS Microbiol. Lett. 125:287-292(1995).  
 CC RN [3]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=CA4098;  
 CC RA Yang S., Huang X.D., Huang Y.H., Li S.Y., Yuan Z.Y.;  
 CC RT "Bacillus megaterium CA4098 penicillin G amidase structural gene  
 CC RT sequence";  
 CC RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 CC CC -1- CATALYTIC ACTIVITY: Penicillin + H(2)O = a fatty acid anion + 6-  
 CC aminopenicillanate.

CC -1- SUBUNIT: HETERODIMER OF AN ALPHA CHAIN AND A BETA CHAIN  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Extracellular (Potential).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S45.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U07682; AAB41343.1; -  
 CC DR EMBL; Z37542; CAAB5774.1; -  
 CC DR EMBL; AF161313; AAD45609.1; -  
 CC DR HSSP; P06875; 1AJQ.  
 CC DR MEROPS; S45.001;  
 CC InterPro: IPR002692; Penicill\_amidase.  
 CC DR Pfam; PF01804; Penicill\_amidase; 1.  
 CC KW Hydrolase; Antibiotic resistance; Zymogen; Signal.  
 CC SIGNAL 1 24 POTENTIAL.  
 CC CHAIN 25 802 PENICILLIN G ACYLASE ZYMOGEN.  
 CC CHAIN 25 234 PENICILLIN G ACYLASE ALPHA SUBUNIT.  
 CC PROPEP 235 265 SPACER PEPTIDE.  
 CC CHAIN 266 802 PENICILLIN G ACYLASE BETA SUBUNIT.  
 CC ACT\_SITE 266 266 BY SIMILARITY.  
 CC SEQUENCE 802 AA; 91987 MW; 877CA0564E50DFBD CRC64;  
 CC  
 CC Query Match 55.6%; Score 35; DB 1; Length 802;  
 CC Best Local Similarity 50.0%; Pred. No. 96;  
 CC Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 CC  
 CC QY 3 EEISEVNXVAEF 14  
 CC |:|:|:| | | |  
 CC Db 548 EDINEINTASF 559  
 CC  
 CC RESULT 15  
 CC RCSA\_KLEAF STANDARD; PRT; 207 AA.  
 CC ID RCSA\_KLEAF P05338;  
 CC AC P05338;  
 CC DT 01-NOV-1988 (Rel. 09, Created)  
 CC DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 CC DT 01-MAR-1992 (Rel. 21, Last annotation update)  
 CC DE Colanic acid capsular biosynthesis activation protein A.  
 CC GN RCSA.  
 CC OS Klebsiella aerogenes.  
 CC OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC OC Klebsiella.  
 CC OX NCBI\_TaxID=28451;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=88009858; PubMed=3309150;  
 CC RA Allen P., Hart C.A., Saunders J.R.;  
 CC RT "Isolation from Klebsiella and characterization of two rcs genes that  
 CC RT activate colanic acid capsular biosynthesis in Escherichia coli";  
 CC RL J. Gen. Microbiol. 133:331-340(1987).  
 CC CC -1- FUNCTION: POSITIVE REGULATOR OF CAPSULAR POLYSACCHARIDE  
 CC SYNTHESIS. RCSA AND RCSB FORM A COMPLEX TO PROMOTE TRANSCRIPTION  
 CC OF THE GENES FOR CAPSULE SYNTHESIS.

CC -!- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M15748; AAA25142.1; -;  
DR PIR; A47615; A47615;  
DR InterPro; IPR000792; HTH\_LuxR.  
DR Pfam; PF00196; GerE; 1.  
DR PRINTS; PR00038; HTHLUXR.  
DR PRODOM; PD000307; HTH\_LuxR; 1.  
DR SMART; SM00421; HTH\_LuxR; 1.  
DR PROSITE; PS00622; HTH\_LUXR\_FAMILY; 1.  
KW Transcription regulation; DNA-binding; Activator.  
FT DNA\_BIND 155 174 H-T-H MOTIF (BY SIMILARITY).  
SQ SEQUENCE 207 AA; 23366 MW; 24E2253BB9DEE162 CRC64;

Query Match 54.0%; Score 34; DB 1; Length 207;  
Best Local Similarity 53.8%; Pred. NO. 36;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAE 13  
Db 26 KKEIVEVNSAAD 38

Search completed: April 23, 2003, 12:35:43  
Job time : 10.1724 secs





; APPLICANT: Jagodzinska, Barbara  
; APPLICANT: John, Varghese  
; APPLICANT: Maillaird, Michel  
; APPLICANT: Pulley, Shon R.  
; APPLICANT: TenBrink, Ruth E.  
; TITLE OF INVENTION: COMPOUNDS TO TREAT ALZHEIMER'S DISEASE  
; FILE REFERENCE: 13615.25USU4  
; CURRENT APPLICATION NUMBER: US/09/896,139  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/215,323  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 60/252,736  
; PRIOR FILING DATE: 2000-11-22  
; PRIOR APPLICATION NUMBER: US 60/255,956  
; PRIOR FILING DATE: 2000-12-15  
; PRIOR APPLICATION NUMBER: US 60/268,497  
; PRIOR FILING DATE: 2001-02-13  
; PRIOR APPLICATION NUMBER: US 60/279,779  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: US 60/295,589  
; PRIOR FILING DATE: 2001-06-04  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 9  
; LENGTH: 30  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-896-139-9

Query Match 85.7%; Score 54; DB 10; Length 30;  
Best Local Similarity 85.7%; Pred. No. 0.0016;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KTEEISEVNXVAEF 14  
|||||  
Db 17 KTEEISEVNLDAEF 30

RESULT 3  
US-09-895-843-9  
; Sequence 9, Application US/09895843  
; Patent No. US20020143177A1  
; GENERAL INFORMATION:  
; APPLICANT: Beck, James P.  
; APPLICANT: Fang, Lawrence Y.  
; APPLICANT: Freskos, John N.  
; APPLICANT: Gailunas, Andrea  
; APPLICANT: Hom, Roy  
; APPLICANT: Jagodzinska, Barbara  
; APPLICANT: John, Varghese  
; APPLICANT: Maillaird, Michel  
; APPLICANT: Pulley, Shon R.  
; APPLICANT: TenBrink, Ruth E.  
; TITLE OF INVENTION: COMPOUNDS TO TREAT ALZHEIMER'S DISEASE  
; FILE REFERENCE: 13615.41USU1  
; CURRENT APPLICATION NUMBER: US/09/895,843  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/215,323  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 9  
; LENGTH: 30  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-895-843-9

Query Match 85.7%; Score 54; DB 10; Length 30;  
Best Local Similarity 85.7%; Pred. No. 0.0016;

Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 KTEEISEVNXVAEF 14  
|||||  
Db 17 KTEEISEVNLDAEF 30  
RESULT 4  
US-09-896-874-4  
; Sequence 4, Application US/09896874  
; Patent No. US20020016320A1  
; GENERAL INFORMATION:  
; APPLICANT: Fang, Lawrence Y.  
; APPLICANT: John, Varghese  
; TITLE OF INVENTION: COMPOUNDS TO TREAT ALZHEIMER'S DISEASE  
; FILE REFERENCE: 13615.40USU1  
; CURRENT APPLICATION NUMBER: US/09/896,874  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/215,323  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 32  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-896-874-4

Query Match 85.7%; Score 54; DB 10; Length 32;  
Best Local Similarity 85.7%; Pred. No. 0.0017;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KTEEISEVNXVAEF 14  
|||||  
Db 17 KTEEISEVNLDAEF 30

RESULT 5  
US-09-896-139-4  
; Sequence 4, Application US/09896139  
; Patent No. US20020128255A1  
; GENERAL INFORMATION:  
; APPLICANT: Beck, James P.  
; APPLICANT: Fang, Lawrence Y.  
; APPLICANT: Freskos, John N.  
; APPLICANT: Gailunas, Andrea  
; APPLICANT: Hom, Roy  
; APPLICANT: Jagodzinska, Barbara  
; APPLICANT: John, Varghese  
; APPLICANT: Maillaird, Michel  
; APPLICANT: Pulley, Shon R.  
; APPLICANT: TenBrink, Ruth E.  
; TITLE OF INVENTION: COMPOUNDS TO TREAT ALZHEIMER'S DISEASE  
; FILE REFERENCE: 13615.25USU4  
; CURRENT APPLICATION NUMBER: US/09/896,139  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/215,323  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 60/252,736  
; PRIOR FILING DATE: 2000-11-22  
; PRIOR APPLICATION NUMBER: US 60/255,956  
; PRIOR FILING DATE: 2000-12-15  
; PRIOR APPLICATION NUMBER: US 60/268,497  
; PRIOR FILING DATE: 2001-02-13  
; PRIOR APPLICATION NUMBER: US 60/279,779  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: US 60/295,589  
; PRIOR FILING DATE: 2001-06-04  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4

; LENGTH: 32  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-896-139-4

Query Match 85.7%; Score 54; DB 10; Length 32;  
Best Local Similarity 85.7%; Pred. No. 0.0017; 2; Indels  
Matches 12; Conservative 0; Mismatches 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
| | | | | | | | | |  
Db 17 KTEEISEVNLDAEF 30

RESULT 6  
US-09-895-843-4  
; Sequence 4, Application US/09895843  
; Patent No. US20020143177A1  
; GENERAL INFORMATION:  
; APPLICANT: Beck, James P.  
; APPLICANT: Fang, Lawrence Y.  
; APPLICANT: Freskos, John N.  
; APPLICANT: Gailunas, Andrea  
; APPLICANT: Hom, Roy  
; APPLICANT: Jagodzinska, Barbara  
; APPLICANT: John, Varghese  
; APPLICANT: Maillaird, Michel  
; APPLICANT: Pulley, Shon R.  
; APPLICANT: TenBrink, Ruth E.  
; TITLE OF INVENTION: COMPOUNDS TO TREAT ALZHEIMER'S DISEASE  
; FILE REFERENCE: 13615.41USU1  
; CURRENT APPLICATION NUMBER: US/09/895,843  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/215,323  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 32  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-895-843-4

Query Match 85.7%; Score 54; DB 10; Length 32;  
Best Local Similarity 85.7%; Pred. No. 0.0017;  
Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
| | | | | | | | | |  
Db 17 KTEEISEVNLDAEF 30

RESULT 7  
US-09-896-874-6  
; Sequence 6, Application US/09896874  
; Patent No. US20020016320A1  
; GENERAL INFORMATION:  
; APPLICANT: Fang, Lawrence Y.  
; APPLICANT: John, Varghese  
; TITLE OF INVENTION: COMPOUNDS TO TREAT ALZHEIMER'S DISEASE  
; FILE REFERENCE: 13615.40USU1  
; CURRENT APPLICATION NUMBER: US/09/896,874  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/215,323  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 33

; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-896-874-6

Query Match 85.7%; Score 54; DB 10; Length 33;  
Best Local Similarity 85.7%; Pred. No. 0.0018;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
| | | | | | | | | |  
Db 20 KTEEISEVNLDAEF 33

RESULT 8  
US-09-896-139-6  
; Sequence 6, Application US/09896139  
; Patent No. US20020128255A1  
; GENERAL INFORMATION:  
; APPLICANT: Beck, James P.  
; APPLICANT: Fang, Lawrence Y.  
; APPLICANT: Freskos, John N.  
; APPLICANT: Gailunas, Andrea  
; APPLICANT: Hom, Roy  
; APPLICANT: Jagodzinska, Barbara  
; APPLICANT: John, Varghese  
; APPLICANT: Maillaird, Michel  
; APPLICANT: Pulley, Shon R.  
; APPLICANT: TenBrink, Ruth E.  
; TITLE OF INVENTION: COMPOUNDS TO TREAT ALZHEIMER'S DISEASE  
; FILE REFERENCE: 13615.25USU4  
; CURRENT APPLICATION NUMBER: US/09/896,139  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/215,323  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 60/252,736  
; PRIOR FILING DATE: 2000-11-22  
; PRIOR APPLICATION NUMBER: US 60/255,956  
; PRIOR FILING DATE: 2000-12-15  
; PRIOR APPLICATION NUMBER: US 60/268,497  
; PRIOR FILING DATE: 2001-02-13  
; PRIOR APPLICATION NUMBER: US 60/279,779  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: US 60/295,589  
; PRIOR FILING DATE: 2001-06-04  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 33  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-896-139-6

Query Match 85.7%; Score 54; DB 10; Length 33;  
Best Local Similarity 85.7%; Pred. No. 0.0018;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
| | | | | | | | | |  
Db 20 KTEEISEVNLDAEF 33

RESULT 9  
US-09-895-843-6  
; Sequence 6, Application US/09895843  
; Patent No. US20020143177A1  
; GENERAL INFORMATION:  
; APPLICANT: Beck, James P.  
; APPLICANT: Fang, Lawrence Y.  
; APPLICANT: Freskos, John N.

APPLICANT: Gailunas, Andrea  
APPLICANT: Hom, Roy  
APPLICANT: Jagodzinska, Barbara  
APPLICANT: John, Varghese  
APPLICANT: Maillard, Michel  
APPLICANT: Pulley, Shon R.  
APPLICANT: TenBrink, Ruth E.  
TITLE OF INVENTION: COMPOUNDS TO TREAT ALZHEIMER'S DISEASE  
FILE REFERENCE: 13615.41USU1  
CURRENT APPLICATION NUMBER: US/09/895,843  
CURRENT FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/215,323  
PRIOR FILING DATE: 2000-06-30  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 6  
LENGTH: 33  
TYPE: PRT  
ORGANISM: Artificial Sequence  
OTHER INFORMATION: Synthetic  
US-09-895-843-6

Query Match 85.7%; Score 54; DB 10; Length 33;  
Best Local Similarity 85.7%; Pred. No. 0.0018;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
DB 20 KTEEISEVNLDAEF 33

RESULT 10  
US-09-794-927-12  
Sequence 12, Application US/09794927  
Patent No. US20010016324A1  
GENERAL INFORMATION:

APPLICANT: Gurney, Mark E.  
APPLICANT: Bienkowski, Michael J.  
APPLICANT: Heinrichson, Robert L.  
APPLICANT: Parodi, Luis A.  
APPLICANT: Yan, Riqiang

TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND  
TITLE OF INVENTION: USES  
TITLE OF INVENTION: THEREFOR

FILE REFERENCE: 28341/6280FG  
CURRENT APPLICATION NUMBER: US/09/794,927  
CURRENT FILING DATE: 2001-02-27  
PRIOR APPLICATION NUMBER: 09/416,901  
PRIOR FILING DATE: 1999-10-13  
PRIOR APPLICATION NUMBER: 60/155,493  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: 09/404,133  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: PCT/US99/20881  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: 60/101,594  
PRIOR FILING DATE: 1998-09-24  
NUMBER OF SEQ ID NOS: 73

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 12  
LENGTH: 695  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-794-927-12

Query Match 85.7%; Score 54; DB 10; Length 695;  
Best Local Similarity 85.7%; Pred. No. 0.051;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
DB 587 KTEEISEVNLDAEF 600

RESULT 11  
US-09-795-847-12  
Sequence 12, Application US/09795847  
Patent No. US20010018208A1  
GENERAL INFORMATION:

APPLICANT: Gurney, Mark E.  
APPLICANT: Bienkowski, Michael J.  
APPLICANT: Heinrichson, Robert L.  
APPLICANT: Parodi, Luis A.

APPLICANT: Yan, Riqiang  
TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR,  
TITLE OF INVENTION: USES  
TITLE OF INVENTION: THEREFOR  
FILE REFERENCE: 28341/6280DE  
CURRENT APPLICATION NUMBER: US/09/795,847  
CURRENT FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: 09/416,901  
PRIOR FILING DATE: 1999-10-13  
PRIOR APPLICATION NUMBER: 60/155,493  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: 09/404,133  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: PCT/US99/20881  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: 60/101,594  
PRIOR FILING DATE: 1998-09-24  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 12  
LENGTH: 695  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-795-847-12

Query Match 85.7%; Score 54; DB 10; Length 695;  
Best Local Similarity 85.7%; Pred. No. 0.051;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
DB 587 KTEEISEVNLDAEF 600

RESULT 12  
US-09-794-743-12  
Sequence 12, Application US/09794743  
Patent No. US20010021391A1  
GENERAL INFORMATION:

APPLICANT: Gurney, Mark E.  
APPLICANT: Bienkowski, Michael J.  
APPLICANT: Heinrichson, Robert L.  
APPLICANT: Parodi, Luis A.

APPLICANT: Yan, Riqiang  
TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR,  
TITLE OF INVENTION: USES  
TITLE OF INVENTION: THEREFOR  
FILE REFERENCE: 28341/6280BC  
CURRENT APPLICATION NUMBER: US/09/794,743  
CURRENT FILING DATE: 2001-02-27  
PRIOR APPLICATION NUMBER: 09/416,901  
PRIOR FILING DATE: 1999-10-13  
PRIOR APPLICATION NUMBER: 60/155,493  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: 09/404,133  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: PCT/US99/20881  
PRIOR FILING DATE: 1999-09-23  
PRIOR APPLICATION NUMBER: 60/101,594  
PRIOR FILING DATE: 1998-09-24  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 12  
; LENGTH: 695  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-794-743-12

Query Match 85.7%; Score 54; DB 10; Length 695;  
Best Local Similarity 85.7%; Pred. No. 0.051;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
||||||| |||  
Db 587 KTEISEVNLD AEF 600

RESULT 13

US-09-794-748-12  
; Sequence 12, Application US/09794748  
; Patent No. US20020037315A1  
; GENERAL INFORMATION:  
; APPLICANT: Gurney, Mark E.  
; APPLICANT: Bienkowski, Michael J.  
; APPLICANT: Heinrikson, Robert L.  
; APPLICANT: Parodi, Luis A.  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND  
; FILE REFERENCE: 28341/6280JL  
; CURRENT APPLICATION NUMBER: US/09/794,748  
; CURRENT FILING DATE: 2001-02-27  
; PRIOR APPLICATION NUMBER: 09/416,901  
; PRIOR FILING DATE: 1999-10-13  
; PRIOR APPLICATION NUMBER: 60/155,493  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 09/404,133  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: PCT/US99/20881  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 60/101,594  
; PRIOR FILING DATE: 1998-09-24  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 695  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-794-748-12

Query Match 85.7%; Score 54; DB 10; Length 695;  
Best Local Similarity 85.7%; Pred. No. 0.051;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
||||||| |||  
Db 587 KTEISEVNLD AEF 600

RESULT 14

US-09-794-925-12  
; Sequence 12, Application US/09794925  
; Patent No. US20020064819A1  
; GENERAL INFORMATION:  
; APPLICANT: Gurney, Mark E.  
; APPLICANT: Bienkowski, Michael J.  
; APPLICANT: Heinrikson, Robert L.  
; APPLICANT: Parodi, Luis A.  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND USES

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND USES  
; FILE REFERENCE: 28341/6280HI  
; CURRENT APPLICATION NUMBER: US/09/794,925  
; CURRENT FILING DATE: 2001-02-27

; PRIOR APPLICATION NUMBER: 09/416,901  
; PRIOR FILING DATE: 1999-10-13  
; PRIOR APPLICATION NUMBER: 60/155,493  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 09/404,133  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: PCT/US99/20881  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 60/101,594  
; PRIOR FILING DATE: 1998-09-24  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 695  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-794-925-12

Query Match 85.7%; Score 54; DB 10; Length 695;  
Best Local Similarity 85.7%; Pred. No. 0.051;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
||||||| |||  
Db 587 KTEISEVNLD AEF 600

RESULT 15

US-09-681-442-12  
; Sequence 12, Application US/09681442  
; Patent No. US20020081634A1  
; GENERAL INFORMATION:  
; APPLICANT: Gurney, Mark E.  
; APPLICANT: Bienkowski, Michael J.  
; APPLICANT: Heinrikson, Robert L.  
; APPLICANT: Parodi, Luis A.  
; APPLICANT: Yan, Riglang  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND USES  
; FILE REFERENCE: 28341/6280FG  
; CURRENT APPLICATION NUMBER: US/09/681,442  
; CURRENT FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 09/416,901  
; PRIOR FILING DATE: 1999-10-13  
; PRIOR APPLICATION NUMBER: 60/155,493  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 09/404,133  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: PCT/US99/20881  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 60/101,594  
; PRIOR FILING DATE: 1998-09-24  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 695  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-681-442-12

Query Match 85.7%; Score 54; DB 10; Length 695;  
Best Local Similarity 85.7%; Pred. No. 0.051;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
||||||| |||  
Db 587 KTEISEVNLD AEF 600

Search completed: April 23, 2003, 12:49:49  
Job time : 29.931 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:33:40 ; Search time 16.8966 Seconds  
(without alignments)  
79.654 Million cell updates/sec

Title: US-09-723-722A-72

Perfect score: 63

Sequence: 1 KTEEISEVNXVAEF 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	48	76.2	82	P00438	Alzheimer's disease
2	48	76.2	695	A49795	Alzheimer's disease
3	48	76.2	695	A27485	Alzheimer's disease
4	48	76.2	695	S00550	Alzheimer's disease
5	48	76.2	770	1 QRHUA4	Alzheimer's disease
6	42	66.7	747	2 JH0773	Alzheimer's disease
7	39	61.9	455	2 S50725	hypothetical prote
8	38	60.3	316	2 T23930	hypothetical prote
9	38	60.3	335	2 A24785	hypothetical prote
10	38	60.3	1354	2 T13930	tripeptidyl-peptid
11	37	58.7	1185	2 A42404	collagen adhesin -
12	36.5	57.9	587	2 T48582	auxin-regulated pr
13	36	57.1	512	2 G95555	55 kDa B regulator
14	36	57.1	513	2 S55889	protein phosphatas
15	36	57.1	650	2 T22002	hypothetical prote
16	36	57.1	837	1 A56681	penicillin amidase
17	36	57.1	837	2 JN0292	antigen 332 - mala
18	36	57.1	867	2 T05389	hypothetical prote
19	35	55.6	184	2 D72745	hypothetical prote
20	35	55.6	351	2 JU0056	flagellin - Serrat
21	35	55.6	433	2 C84619	probable serine ca
22	35	55.6	456	2 S19212	adenylosuccinate i
23	35	55.6	456	2 A85687	adenylosuccinate i
24	35	55.6	456	2 C90829	adenylosuccinate i
25	35	55.6	504	2 AD1615	glycine betaine/ca
26	35	55.6	505	2 E92446	4-alpha-glucanotra
27	35	55.6	505	2 B98111	4-alpha-glucanotra
28	35	55.6	572	2 S72249	trithorax protein
29	35	55.6	731	2 T19721	hypothetical prote

30 35 55.6 802 2 S49252 penicillin amidase  
31 35 55.6 802 2 I39665 penicillin amidase  
32 34 54.0 136 2 C81436 biopolymer transpo  
33 34 54.0 207 2 A47615 colanic acid capsu  
34 34 54.0 225 2 E97244 cotJC-like protein  
35 34 54.0 274 2 B96991 methyl-accepting c  
36 34 54.0 294 2 D97693 citrate lyase, bet  
37 34 54.0 294 2 AC2919 citrate lyase, bet  
38 34 54.0 307 2 S38152 mrp protein (impor  
39 34 54.0 319 2 C97720 mrp protein (impor  
40 34 54.0 341 2 F71042 hypothetical prote  
41 34 54.0 352 2 C75172 enolase related (e  
42 34 54.0 357 2 E71123 probable sugar tra  
43 34 54.0 366 2 B87446 hypothetical prote  
44 34 54.0 412 2 C38351 phosphoprotein pho  
45 34 54.0 415 2 T12974 hypothetical prote

#### ALIGNMENTS

##### RESULT 1

PQ0438

Alzheimer's disease amyloid A4 protein precursor - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 30-Sep-1993 #sequence\_revision 19-Oct-1995 #text\_change 19-Oct-1995

C:Accession: PQ0438; C60045

R:Davidson, J.S.; West, R.L.; Kotikalapudi, P.; Maroun, L.E.

Biochem. Biophys. Res. Commun. 188, 905-911, 1992

A:Title: Sequence and methylation in the beta/A4 region of the rabbit amyloid precurs

A:Reference number: PQ0438; MUID:93075180; PMID:1445331

A:Accession: PQ0438

A:Molecule type: DNA

A:Residues: 1-82 <DAV>

A:Cross-references: GB:M83558; GB:M83657

R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.

Brain Res. Mol. Brain Res. 10, 299-305, 1991

A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d

A:Reference number: A60045; MUID:92017079; PMID:1656157

A:Accession: C60045

A:Molecule type: mRNA

A:Residues: 12-68 <JOH>

A:Cross-references: EMBL:X56129

C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas

C:Keywords: alternative splicing; Alzheimer's disease; amyloid; Down's syndrome

Query Match 76.2%; Score 48; DB 2; Length 82;

Best Local Similarity 78.6%; Pred. No. 0.054;

Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14

Db 7 KTEEISEVKNDAEF 20

##### RESULT 2

A49795

Alzheimer's disease amyloid beta protein precursor - crab-eating macaque

C:Species: Macaca fascicularis (crab-eating macaque)

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: A49795

R:Podlisky, M.B.; Tolan, D.R.; Selkoe, D.J.

Am. J. Pathol. 138, 1423-1435, 1991

A:Title: Homology of the amyloid beta protein precursor in monkey and human supports

A:Reference number: A49795; MUID:91273117; PMID:1905108

A:Accession: A49795

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-695 <POD>

A:Cross-references: GB:IM58727; MUID:9342062; PIDN:AAA36829.1; PID:9342063

C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas

C:Keywords: alternative splicing

Query Match 76.2% Score 48; DB 1; Length 695;  
Best Local Similarity 78.6%; Pred. No. 0.58;  
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
|||||  
Db 587 KTEEISEVKMDAEF 600

RESULT 3  
A27485  
Alzheimer's disease amyloid beta/A4 protein homolog precursor - mouse  
N:Alternate names: proteinase nexin II  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 13-Aug-1999  
C:Accession: A27485; S19727; I49485  
R:Yamada, T.; Sakaki, H.; Furuya, H.; Miyata, T.; Goto, I.; Sakaki, Y.  
Biochem. Biophys. Res. Commun. 149, 665-671, 1987  
A:Title: Complementary DNA for the mouse homolog of the human amyloid beta protein precursor  
A:Reference number: A27485; MUID:88106489; PMID:3322280  
A:Accession: A27485  
A:Molecule type: mRNA  
A:Residues: 1-695 <FAM>  
A:Cross-references: GB:M18373; NID:g191568; PIDN:AAA37139.1; PID:g309085  
A:Experimental source: brain  
R:de Strooper, B.; van Leuven, F.; van den Berghe, H.  
Biochim. Biophys. Acta 1129, 141-143, 1991  
A:Title: The amyloid beta protein precursor or proteinase nexin II from mouse is closer  
A:Reference number: S19727; MUID:92096458; PMID:I756177  
A:Accession: S19727  
A:Molecule type: mRNA  
A:Residues: 1-210, 'G', 212-220, 'S', 222-396, 'A', 398-402, 'T', 404-448, 'A', 450-695 <STR>  
A:Cross-references: EMBL:X59379  
R:Izumi, R.; Yamada, T.; Yoshikai, S.; Sakaki, H.; Hattori, M.; Sakaki, Y.  
Gene 112, 189-195, 1992  
A:Title: Positive and negative regulatory elements for the expression of the Alzheimer's  
A:Reference number: I49485; MUID:92209998; PMID:1555768  
A:Accession: I49485  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-19 <RES>  
A:Cross-references: GB:D10603; NID:g220328; PIDN:BAA01456.1; PID:g220329  
C:Genetics:  
C:Map position: 16C3  
C:Superfamily: Alzheimer's disease amyloid beta protein; animal kunitz-type proteinase  
C:Keywords: alternative splicing; amyloid; transmembrane protein

Query Match 76.2% Score 48; DB 2; Length 695;  
Best Local Similarity 78.6%; Pred. No. 0.58;  
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
|||||  
Db 587 KTEEISEVKMDAEF 600

RESULT 4  
S00550  
Alzheimer's disease amyloid beta protein precursor - rat  
N:Alternate names: beta-A4 amyloid protein  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 13-Aug-1999  
C:Accession: S00550; A41245; A39820; S46251  
R:Shivers, B.D.; Hilbich, C.; Multhaup, G.; Salbaum, M.; Beyreuther, K.; Seeburg, P.H.  
EMBO J. 7, 1365-1370, 1988  
A:Title: Alzheimer's disease amyloidogenic glycoprotein: expression pattern in rat brain  
A:Reference number: S00550; MUID:88312583; PMID:2900758  
A:Accession: S00550  
A:Molecule type: mRNA  
A:Residues: 1-695 <SHI>  
A:Cross-references: EMBL:X07648; NID:g55616; PIDN:CAA30488.1; PID:g55617  
R:Schubert, D.; Schroeder, R.; Lacorbiere, M.; Saitoh, T.; Cole, G.  
Science 241, 223-226, 1988

A:Title: Amyloid beta protein precursor is possibly a heparan sulfate proteoglycan  
A:Reference number: A41245; MUID:88264430; PMID:2968652  
A:Accession: A41245  
A:Molecule type: protein  
A:Residues: 18-37, 'X', 39-40, 'X', 42-44 <SCH>  
A:Note: evidence for heparan sulfate attachment  
R:Hesse, L.; Beher, D.; Masters, C.L.; Multhaup, G.  
FEBS Lett. 349, 109-116, 1994  
A:Title: The beta-A4 amyloid precursor protein binding to copper.  
A:Reference number: S46251; MUID:94320627; PMID:7913895  
A:Contents: annotation: copper binding sites  
A:Note: rat peptides were isolated but not sequenced  
R:Potempska, A.; Styles, J.; Mehta, P.; Kim, K.S.; Miller, D.L.  
J. Biol. Chem. 266, 8464-8469, 1991  
A:Title: Purification and tissue level of the beta-amyloid peptide precursor of  
A:Reference number: A39820; MUID:91217087; PMID:1673681  
A:Accession: A39820  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 18-32 <POT>  
A:Experimental source: brain  
C:Comment: Deposition of amyloid protein as neurofibrillary tangles and/or plaques  
C:Superfamily: Alzheimer's disease amyloid beta protein; animal kunitz-type protein  
C:Keywords: alternative splicing; amyloid; glycoprotein; transmembrane protein  
F:625-648/Domain: transmembrane #status predicted <TMM>

Query Match 76.2% Score 48; DB 2; Length 695;  
Best Local Similarity 78.6%; Pred. No. 0.58;  
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
|||||  
Db 587 KTEEISEVKMDAEF 600

RESULT 5  
QRH0A4  
Alzheimer's disease amyloid beta protein precursor [validated] - human  
N:Alternate names: Alzheimer's disease amyloid A4 protein; coagulation factor XI  
N:Contents: amyloid beta protein long, plaque form; amyloid beta protein short,  
protein precursor splice form APP(770)  
C:Species: Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 28-Jul-1995 #text\_change 15-Sep-2000  
C:Accession: S02260; S05194; A32277; A33260; A35486; I39452; I39451; I39453; I5  
4668; A28583; A29302; A60805; J10038; S06121; A60355; A59011; A38384; S29076; S  
R:Lemaire, H.G.; Salbaum, J.M.; Multhaup, G.; Kang, J.; Bayney, R.M.; Unterbeck  
Nucleic Acids Res. 17, 517-522, 1989  
A:Title: The PreA4(695) precursor protein of Alzheimer's disease A4 amyloid is  
A:Reference number: S02260; MUID:89128427; PMID:2783775  
A:Accession: S02260  
A:Molecule type: DNA  
A:Residues: 1-288, 'V', 365-770 <LEM1>  
A:Cross-references: EMBL:X13466  
A:Note: alternative splice form APP(695)  
R:Lemaire, H.G.  
submitted to the EMBL Data Library, November 1988  
A:Reference number: S05194  
A:Accession: S05194  
A:Molecule type: DNA  
A:Residues: 1-14, 'VW', 17-288, 'V', 365-770 <LEM2>  
A:Cross-references: EMBL:X13466; NID:g35598; PIDN:CAA31830.1; PID:g871360  
A:Note: alternative splice form APP(695)  
R:La Fauce, G.; Lahiri, D.K.; Salton, S.R.J.; Robakis, N.K.  
Biochem. Biophys. Res. Commun. 159, 297-304, 1989  
A:Title: Characterization of the 5'-end region and the first two exons of the b  
A:Reference number: A32277; MUID:89165870; PMID:2538123  
A:Accession: A32277  
A:Molecule type: DNA  
A:Residues: 1-75 <LAF>  
A:Cross-references: GB:M24546; GB:M24547; NID:g341202; PIDN:AAC13654.1; PID:g51  
R:Johnstone, E.M.; Chaney, M.O.; Moore, R.E.; Ward, K.E.; Norris, F.H.; Little  
Biochem. Biophys. Res. Commun. 163, 1248-1255, 1989  
A:Title: Alzheimer's disease amyloid peptide is encoded by two exons and shows



A:Reference number: A33260; MUID:89392030; PMID:2675837  
 A:Accession: A33260  
 A:Molecule type: DNA  
 A:Residues: 656-737 <JOH>  
 A:Cross-references: GB:M29270; NID:gl178863; PIDN:AA51768.1; PID:gl178865  
 R:Prelli, F.; Levy, E.; van Duinen, S.G.; Bots, G.T.A.M.; Luyendijk, W.; Frangione, B.  
 Biochem. Biophys. Res. Commun. 170, 301-307, 1990  
 A:Title: Expression of a normal and variant Alzheimer's beta-protein gene in amyloid of  
 A:Reference number: A35486; MUID:90321244; PMID:2196878  
 A:Accession: A35486  
 A:Molecule type: DNA  
 A:Residues: 672-710 <PRE1>  
 A:Note: 693-Gln was found in DNA isolated from HCHWA-D patients  
 R:Yoshikai, S.I.; Sasaki, H.; Doh-ura, K.; Furuya, H.; Sakaki, Y.  
 Gene 87, 257-263, 1990  
 A:Title: Genomic organization of the human amyloid beta-protein precursor gene.  
 A:Reference number: I39451; MUID:90236318; PMID:2110105  
 A:Accession: I39452  
 A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL  
 A:Molecule type: DNA  
 A:Residues: 1-770 <VOS1>  
 A:Cross-references: GB:M33112; NID:gl178613; PIDN:AAB59502.1; PID:gl178616  
 A:Accession: I39451  
 A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL  
 A:Molecule type: DNA  
 A:Residues: 1-530, 'QWLMPVIPAFWEAKVGR' <VOS2>  
 A:Cross-references: GB:M34875; NID:gl178608; PIDN:AAB59501.1; PID:gl178615  
 R:Yoshikai, S.I.; Sasaki, H.; Doh-ura, K.; Furuya, H.; Sakaki, Y.  
 Gene 102, 291-292, 1991  
 A:Reference number: A59020; MUID:91340168; PMID:1908403  
 A:Contents: annotation; erratum  
 A:Note: revised physical map for reference I39451  
 R:Levy, E.; Carman, M.D.; Fernandez-Madrid, I.J.; Power, M.D.; Lieberburg, I.; van Duinen  
 Science 248, 1124-1126, 1990  
 A:Title: Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrh  
 A:Reference number: I39453; MUID:90260863; PMID:2111584  
 A:Accession: I39453  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 656-737 <LEV>  
 A:Cross-references: GB:M37896; NID:gl178618; PIDN:AA51727.1; PID:gl178620  
 A:Note: a mutation with 693-Gln is presented  
 R:Murrell, J.; Farlow, M.; Ghetti, B.; Benson, M.D.  
 Science 254, 97-99, 1991  
 A:Title: A mutation in the amyloid precursor protein associated with hereditary Alzheimer  
 A:Reference number: I59562; MUID:92022553; PMID:1925564  
 A:Accession: I59562  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 689-716, 'F', 718-737 <MUR>  
 A:Cross-references: GB:S57665; NID:q236720; PIDN:AAB19991.1; PID:q236721  
 R:Kamino, K.; Orr, H.T.; Payami, H.; Wijsman, E.M.; Alonso, M.E.; Pulst, S.M.; Anderson,  
 arakis, S.E.; Korenberg, J.R.; Sharma, V.; Kukull, W.; Larson, E.; Heston, L.L.; Martin,  
 Am. J. Hum. Genet. 51, 998-1014, 1992  
 A:Title: Linkage and mutational analysis of familial Alzheimer disease kindreds for the  
 A:Reference number: A44017; MUID:93035397; PMID:1415269  
 A:Accession: A44017  
 A:Molecule type: DNA  
 A:Residues: 687-692, 'G', 694-718 <KAM1>  
 A:Cross-references: GB:S45135; NID:q257377; PIDN:AAB23645.1; PID:q257378  
 A:Experimental source: familial Alzheimer disease family SB  
 A:Note: sequence extracted from NCBI backbone (NCBIP:115374)  
 A:Accession: B44017  
 A:Molecule type: DNA  
 A:Residues: 687-718 <RAM2>  
 A:Cross-references: GB:S45136; NID:q257379; PIDN:AAB23646.1; PID:q257380  
 A:Experimental source: familial Alzheimer disease family LIT  
 A:Note: sequence extracted from NCBI backbone (NCBIP:115376)  
 A:Title: this sequence has a silent mutation  
 R:Kang, J.; Lemaire, H.G.; Unterbeck, A.; Salbaum, J.M.; Masters, C.L.; Grzeschik, K.H.;  
 Nature 325, 733-736, 1987  
 A:Title: The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surfac  
 A:Reference number: A03134; MUID:87144572; PMID:2881207  
 A:Accession: A03134  
 A:Molecule type: mRNA  
 A:Residues: 1-288, 'V', 365-770 <KAN>  
 A:Cross-references: GB:Y00264; NID:q28525; PIDN:CAA68374.1; PID:q28526  
 A:Note: alternative splice form APP(695)  
 R:Robakis, N.K.; Ramakrishna, N.; Wolfe, G.; Wisniewski, H.M.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 4190-4194, 1987  
 A:Title: Molecular cloning and characterization of a cDNA encoding the cerebrovascula  
 A:Reference number: A29030; MUID:87231971; PMID:3035574  
 A:Accession: A29030  
 A:Molecule type: mRNA  
 A:Residues: 284-288, 'V', 365-646, 'E', 648-770 <ROB>  
 A:Cross-references: GB:M16765; NID:gl178539; PIDN:AAA51722.1; PID:gl178540  
 R:Goldgaber, D.; Lerman, M.I.; McBride, O.W.; Saffiotti, U.; Gajdusek, D.C.  
 Science 235, 877-880, 1987  
 A:Title: Characterization and chromosomal localization of a cDNA encoding brain amylo  
 A:Reference number: A47584; MUID:87120328; PMID:3810169  
 A:Accession: A47584  
 A:Molecule type: mRNA  
 A:Residues: 674-756, 'S', 758-770 <GOL>  
 A:Cross-references: GB:M15533; NID:gl178706; PIDN:AAA35540.1; PID:gl178707  
 A:Experimental source: brain  
 R:Tanzi, R.E.; Gusella, J.F.; Watkins, P.C.; Bruns, G.A.P.; St George-Hyslop, P.; Van  
 Science 235, 880-884, 1987  
 A:Title: Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near  
 A:Reference number: A47585; MUID:87120329; PMID:2949367  
 A:Accession: A47585  
 A:Molecule type: mRNA  
 A:Residues: 674-703 <TAN1>  
 A:Cross-references: GB:M15532; NID:gl177957; PIDN:AAA51564.1; PID:gl177958  
 R:Dyck, T.; Weidemann, A.; Multhaup, G.; Salbaum, J.M.; Lemaire, H.G.; Kang, J.; Mue  
 EMBO J. 7, 949-957, 1988  
 A:Title: Identification, transmembrane orientation and biogenesis of the amyloid A4 p  
 A:Reference number: S02638; MUID:88296437; PMID:2900137  
 A:Accession: S02638  
 A:Molecule type: mRNA  
 A:Residues: 672-678 <DYR>  
 R:Tanzi, R.E.; McClatchey, A.I.; Lamperti, E.D.; Villa-Komaroff, L.; Gusella, J.F.; M  
 Nature 331, 528-530, 1988  
 A:Title: Protease inhibitor domain encoded by an amyloid protein precursor mRNA assoc  
 A:Reference number: S00707; MUID:88122640; PMID:2893290  
 A:Accession: S00707  
 A:Molecule type: mRNA  
 A:Residues: 286-344, 'I', 365-366 <TAN2>  
 A:Cross-references: EMBL:X06982; NID:q28817; PIDN:CRAA30042.1; PID:q929612  
 A:Experimental source: promyelocytic leukemia cell line HL60  
 A:Note: alternative splice form APP(751)  
 R:Ponte, P.; Gonzalez-DeWhitt, P.; Schilling, J.; Miller, J.; Hsu, D.; Greenberg, B.;  
 Nature 331, 525-527, 1988  
 A:Title: A new A4 amyloid mRNA contains a domain homologous to serine proteinase inh  
 A:Reference number: S00925; MUID:88122639; PMID:2893289  
 A:Accession: S00925  
 A:Molecule type: mRNA  
 A:Residues: 1-344, 'I', 365-770 <PO2>  
 A:Cross-references: GB:X06989; EMBL:Y00297; NID:q28720; PIDN:CAA30050.1; PID:q28721  
 A:Note: alternative splice form APP(751)  
 R:Kitaguchi, N.; Takahashi, Y.; Tokushima, Y.; Shiojiri, S.; Ito, H.  
 Nature 331, 530-532, 1988  
 A:Title: Novel precursor of Alzheimer's disease amyloid protein shows protease inhibi  
 A:Reference number: A38949; MUID:88122641; PMID:2893291  
 A:Accession: A38949  
 A:Molecule type: mRNA  
 A:Residues: 287-367 <KIT>  
 A:Cross-references: GB:X06981; NID:q28816; PIDN:CAA30041.1; PID:q929611  
 A:Experimental source: glioblastoma cell line  
 A:Note: alternative splice form APP(770)  
 R:Vitek, M.P.; Rasool, C.G.; de Sauvage, F.; Vitek, S.M.; Bartus, R.T.; Beer, B.; Ash  
 Brain Res. Mol. Brain Res. 4, 121-131, 1988  
 A:Title: Absence of mutation in the beta-amyloid cDNAs cloned from the brains of thre  
 A:Reference number: A30320  
 A:Accession: A30320  
 A:Status: not compared with conceptual translation

A:Molecule type: mRNA  
 A:Residues: 284-288, 'V', 365-770 <VIT1>  
 A:Accession: B30320  
 A>Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 122-288, 'V', 365-770 <VIT>  
 A:Accession: C30320  
 A>Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 606-770 <VIT3>  
 R:Zain, S.B.; Salim, M.; Chou, W.G.; Sajdel-Sulkowska, E.M.; Majocha, R.E.; Marotta, C.A.  
 Proc. Natl. Acad. Sci. U.S.A. 85, 929-933, 1988  
 A:Title: Molecular cloning of amyloid cDNA derived from mRNA of the Alzheimer disease br  
 A:Reference number: A31087; MUID:88124954; PMID:2893379  
 A:Accession: A31087  
 A:Molecule type: mRNA  
 A:Residues: 507-770 <ZAI>  
 A:Cross-references: GB:M18734; NID:gl178572; PIDN:AA51726.1; PID:gl178573  
 A:Note: the authors translated the codon GAA for residue 599 as Gly, ACC for residue 603  
 8 as Val, GTG for residue 609 as Asn, AAT for residue 610 as Gly, and GGT for residue 65  
 A:Note: the cited Genbank accession number, J03594, is not in release 101.0  
 R:Masters, C.L.; Multhaup, G.; Simms, G.; Pottgiesser, J.; Martins, R.N.; Beyreuther, K.  
 Query Match 76.2%; Score 48; DB 1; Length 770;  
 Best Local Similarity 78.6%; Pred. No. 0.65;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 KTEEISEVNXVAEF 14  
 Db 662 KTEEISEVKMDAEF 675  
 RESULT 6  
 JH0773  
 Alzheimer's disease amyloid beta protein precursor - African clawed frog  
 C:Species: Xenopus laevis (African clawed frog)  
 C>Date: 10-Jun-1993 #sequence\_revision 10-Jun-1993 #text\_change 13-Aug-1999  
 C:Accession: JH0773  
 R:Okado, H.; Okamoto, H.  
 Biochem. Biophys. Res. Commun. 189, 1561-1568, 1992  
 A:Title: A Xenopus homologue of the human beta-amyloid precursor protein: developmental  
 A:Reference number: JH0773; MUID:93129227; PMID:1282805  
 A:Accession: JH0773  
 A:Molecule type: mRNA  
 A:Residues: 1-747 <OKA>  
 A:Cross-references: GB:S52417; NID:g263150; PIDN:AB24853.1; PID:g263151  
 A:Experimental source: larva  
 C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase i  
 C:Keywords: alternative splicing; amyloid  
 F:287-337/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>  
 Query Match 66.7%; Score 42; DB 2; Length 747;  
 Best Local Similarity 64.3%; Pred. No. 8.7;  
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 KTEEISEVNXVAEF 14  
 Db 639 KTEEISEVKMDSEY 652  
 RESULT 7  
 S50725  
 hypothetical protein YNL206c - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein N1346  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 23-Aug-1995 #sequence\_revision 19-Jan-1996 #text\_change 19-Apr-2002  
 C:Accession: S50725; S63163; S49863  
 R:Jonniaux, J.L.; Coster, F.; Purnelle, B.; Goffeau, A.  
 Yeast 10, 1639-1645, 1994  
 A:Title: A 21.7 kb DNA segment on the left arm of yeast chromosome XIV carries WH13, GCR  
 A:Reference number: S50712; MUID:95242839; PMID:7725799  
 A:Accession: S50725  
 A>Status: nucleic acid sequence not shown

A:Molecule type: DNA  
 A:Residues: 1-455 <JON>  
 A:Cross-references: EMBL:X7898; NID:g600045; PIDN:CAA55502.1; PID:g600059  
 R:Coster, F.; Jonniaux, J.L.; Goffeau, A.; Purnelle, B.; Van Dyck, L.  
 submitted to the Protein Sequence Database, April 1996  
 A:Reference number: S63151  
 A:Accession: S63163  
 A:Molecule type: DNA  
 A:Residues: 1-455 <COS>  
 A:Cross-references: EMBL:Z71482; NID:gl302206; PID:e239612; PID:gl302207; MIPS:Y  
 A:Experimental source: strain S288C  
 C:Genetics:  
 A:Gene: SGD:RTT106  
 A:Cross-references: SGD:S0005150  
 A:Map position: 14L  
 Query Match 61.9%; Score 39; DB 2; Length 455;  
 Best Local Similarity 63.6%; Pred. No. 18;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 KTEEISEVNXV 11  
 Db 63 KTEEISEVTNTI 73  
 RESULT 8  
 T23930  
 hypothetical protein R05D7.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T23930  
 R:Dobson, R.  
 submitted to the EMBL Data Library, October 1996  
 A:Reference number: Z19819  
 A:Accession: T23930  
 A>Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-316 <WIL>  
 A:Cross-references: EMBL:Z81105; PIDN:CAB03217.1; CSPDB:GN00019; CESP:R05D7.3  
 A:Experimental source: clone R05D7  
 C:Genetics:  
 A:Gene: CESP:R05D7.3  
 A:Map position: 1  
 A:Introns: 63/3; 97/3; 122/3; 195/2; 230/1  
 Query Match 60.3%; Score 38; DB 2; Length 316;  
 Best Local Similarity 46.2%; Pred. No. 19;  
 Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 KTEEISEVNXVAE 13  
 Db 283 ETEEITEISEISE 295  
 RESULT 9  
 A24785  
 hypothetical protein 335 - slime mold (Dictyostellium discoideum) transposon DIRS  
 C:Species: Dictyostellium discoideum  
 C>Date: 23-Aug-1987 #sequence\_revision 23-Aug-1987 #text\_change 18-Jun-1993  
 C:Accession: A24785  
 R:Cappello, J.; Handelsman, K.; Lodish, H.F.  
 Cell 43, 105-115, 1985  
 A:Title: Sequence of Dictyostellium DIRS-1: an apparent retrotransposon with inve  
 A:Reference number: A94654; MUID:86079481; PMID:2416457  
 A:Accession: A24785  
 A:Molecule type: DNA  
 A:Residues: 1-335 <CAP>  
 Query Match 60.3%; Score 38; DB 2; Length 335;  
 Best Local Similarity 57.1%; Pred. No. 20;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 KTEEISEVNXVAEF 14

Db 115 KKDEISELNKVFNF 128  
| : : : : : | | |

## RESULT 10

Tl3930  
tripeptidyl-peptidase II (EC 3.4.14.10) - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
C:Accession: Tl3930  
R:Renn, S.C.P.; Tomkinson, B.; Taghert, P.H.  
J. Biol. Chem. 273, 19173-19182, 1998  
A:Title: Characterization and cloning of tripeptidyl-peptidase II from the fruit fly, *D. melanogaster*  
A:Reference number: Z17814; MUID:98334655; PMID:9668104  
A:Accession: Tl3930  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1354 <REN>  
A:Cross-references: EMBL:AF035251; NID:g3387807; PID:g3387808; PIDN:AAC28563.1  
C:Genetics:  
A:Cross-references: FlyBase:FBgn0020370  
A:Map position: 2  
C:Keywords: dipeptidylpeptide hydrolase

Query Match 60.3%; Score 38; DB 2; Length 1354;  
Best Local Similarity 66.7%; Pred. No. 96;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 REISEVNKVAEF 14  
| : : : : | | |

Db 929 EKISEANLVASF 940

## RESULT 11

A42404  
collagen adhesin - *Staphylococcus aureus*  
C:Species: *Staphylococcus aureus*  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 10-Nov-1995  
C:Accession: A42404; S27665  
R:Patti, J.M.; Jonsson, H.; Guss, B.; Switalski, L.M.; Wiberg, K.; Lindberg, M.; Hook, M.  
J. Biol. Chem. 267, 4766-4772, 1992  
A:Title: Molecular characterization and expression of a gene encoding a *Staphylococcus aureus* collagen adhesin  
A:Reference number: A42404; MUID:92165839; PMID:1311320  
A:Contents: FDA 574  
A:Accession: A42404  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1185 <PAT>  
A:Cross-references: EMBL:M81736  
A:Note: sequence extracted from NCBI backbone (NCBIP:83982)

Query Match 58.7%; Score 37; DB 2; Length 1185;  
Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KTEISEVNKVAEF 14  
| : : : : | | |

Db 118 KVEKLSVSGPAEF 131

## RESULT 12

T48582  
auxin-regulated protein GH3 homolog Tl31B5.170 - *Arabidopsis thaliana*  
N:Alternate names: protein Tl31B5.170  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 18-Jul-2001  
C:Accession: T48582  
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, I.  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: 224490  
A:Accession: T48582  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-587 <BEV>  
A:Cross-references: EMBL:AL163491  
A:Experimental source: cultivar Columbia; BAC clone T31B5  
C:Genetics:  
A:Map position: 5  
A:Introns: 97/2; 131/2; 178/2; 400/1  
A:Note: T31B5.170  
C:Superfamily: soybean auxin-regulated protein GH3

Query Match 57.9%; Score 36.5; DB 2; Length 587;  
Best Local Similarity 58.8%; Pred. No. 73;  
Matches 10; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

Qy 1 KTEE---ISEVNKVAEF 14  
| : : : : | | |

Db 562 KTEAVKILEANVSEF 578

## RESULT 13

G96555  
55 kDa B regulatory subunit of phosphatase 2A [imported] - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: G96555  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon, U.; Cohen, N.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzla, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: G96555  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-512 <SNO>  
A:Cross-references: GB:AE005173; NID:gl1094761; PIDN:AAG29694.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 57.1%; Score 36; DB 2; Length 512;  
Best Local Similarity 57.1%; Pred. No. 78;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KTEISEVNKVAEF 14  
| : : : : | | |

Db 277 KMEDLSEVITSAEF 290

## RESULT 14

S55889  
protein phosphatase 2A B regulatory chain 55K - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C:Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 21-Jul-2000  
C:Accession: S55889  
R:Rundle, S.J.; Hartung, A.J.; Corum III, J.W.; O'Neill, M.  
Plant Mol. Biol. 28, 257-266, 1995  
A:Title: Characterization of a cDNA encoding the 55 kDa B regulatory subunit of *Arabidopsis thaliana* protein phosphatase 2A  
A:Reference number: S55889; MUID:95322588; PMID:7599311  
A:Accession: S55889  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-513 <RUN>  
A:Cross-references: GB:U18129; NID:g710329; PIDN:AAA86695.1; PID:g710330

Query Match 57.1%; Score 36; DB 2; Length 513;  
Best Local Similarity 57.1%; Pred. No. 78;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KTEEISEVNXVAEF 14  
| : : : : :  
Db 278 KMEDLSEVITSAEF 291

## RESULT 15

T22002

Hypothetical protein F39H11.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T22002

R:White, S.

Submitted to the EMBL Data Library, October 1996

A:Reference number: Z19500

A:Accession: T22002

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-650 &lt;WIL&gt;

A:Cross-references: EMBL:Z81079; PIDN:CAE03084.1; GSPDB:GN00019; CESP:F39H11.4

A:Experimental source: clone F39H11

C:Genetics:

A:Gene: CESP:F39H11.4

A:Map position: 1

A:Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match

Best Local Similarity 57.1%; Score 36; DB 2; Length 650;

Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KTEEISEVNXVAEF 14  
| : : : : :  
Db 122 KQESSELOKISDF 135

Search completed: April 23, 2003, 12:36:25  
Job time : 18.8966 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2003, 12:35:24 ; Search time 22.6897 seconds  
(without alignments)  
18.155 Million cell updates/sec

Title: US-09-723-722a-72  
Sequence: 63  
1 KTEEISEVNXVAEF 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A-COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B-COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	85.7	30	2	US-08-659-984A-17
2	54	85.7	30	4	US-08-660-531-17
3	54	85.7	33	2	US-08-659-984A-16
4	54	85.7	33	4	US-08-660-531-16
5	54	85.7	42	2	US-08-659-984A-15
6	54	85.7	42	4	US-08-660-531-15
7	54	85.7	506	2	US-08-659-984A-21
8	54	85.7	506	4	US-08-660-531-21
9	54	85.7	506	4	US-09-054-334-4
10	54	85.7	695	4	US-09-548-372D-12
11	54	85.7	695	4	US-09-548-367D-12
12	54	85.7	697	4	US-09-548-372D-18
13	54	85.7	697	4	US-09-548-367D-18
14	48	76.2	58	1	US-08-371-930-25
15	48	76.2	58	5	PCT-US94-01712-25
16	48	76.2	152	6	5187153-4
17	48	76.2	162	6	5220013-4
18	48	76.2	162	6	5223482-4
19	48	76.2	264	1	US-07-990-893-5
20	48	76.2	487	1	US-08-462-859A-9
21	48	76.2	487	1	US-08-123-659A-9
22	48	76.2	487	1	US-08-464-247A-9
23	48	76.2	487	1	US-08-464-248A-9
24	48	76.2	492	1	US-08-462-859A-7
25	48	76.2	492	1	US-08-123-659A-7
26	48	76.2	492	1	US-08-464-247A-7
27	48	76.2	492	1	US-08-464-248A-7

28	48	76.2	656	1	US-08-371-930-23	Sequence 23, Appl
29	48	76.2	656	5	PCT-US94-01712-23	Sequence 23, Appl
30	48	76.2	676	1	US-08-371-930-24	Sequence 24, Appl
31	48	76.2	676	5	PCT-US94-01712-24	Sequence 24, Appl
32	48	76.2	694	1	US-08-339-152A-18	Sequence 18, Appl
33	48	76.2	694	2	US-08-007-999B-5	Sequence 5, Appl
34	48	76.2	694	2	US-08-689-276A-5	Sequence 5, Appl
35	48	76.2	695	1	US-08-371-930-27	Sequence 27, Appl
36	48	76.2	695	1	US-08-123-702-2	Sequence 2, Appl
37	48	76.2	695	1	US-08-339-152A-30	Sequence 30, Appl
38	48	76.2	695	2	US-08-104-165-1	Sequence 1, Appl
39	48	76.2	695	3	US-08-464-250-1	Sequence 1, Appl
40	48	76.2	695	4	US-08-464-250-1	Sequence 1, Appl
41	48	76.2	695	4	US-09-458-481B-4	Sequence 4, Appl
42	48	76.2	695	4	US-09-458-481B-5	Sequence 5, Appl
43	48	76.2	695	4	US-09-458-481B-6	Sequence 6, Appl
44	48	76.2	695	4	US-09-458-481B-7	Sequence 7, Appl
45	48	76.2	695	4	US-09-458-481B-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1  
US-08-659-984A-17  
; Sequence 17, Application US/08659984A  
; Patent No. 5942400  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, John P.  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Jacobson-Croak, Kirsten L.  
; TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
; TITLE OF INVENTION: Inhibition  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/659,984A  
; FILING DATE: 07-JUN-1996  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,152  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 15270-002810US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-659-984A-17

Query Match 85.7%; Score 54; DB 2; Length 30;  
Best Local Similarity 85.7%; Pred. No. 0.00044;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Caps 0;

Qy 1 KTEEISEVNXVAEF 14  
 Db 17 KTEEISEVNLDAEF 30

## RESULT 2

US-08-660-531-17  
 ; Sequence 17, Application US/08660531  
 ; Patent No. 6221645  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chrysler, Susanna M.S.  
 ; APPLICANT: Sinha, Sukanto  
 ; APPLICANT: Keim, Pamela S.  
 ; APPLICANT: Anderson, John P.  
 ; TITLE OF INVENTION: Beta-Secretase  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Townsend and Townsend and Crew LLP  
 ; STREET: Two Embarcadero Ctr., 8th Floor  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94111-3834  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION NUMBER: US/08/660,531  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/480,498  
 ; FILING DATE: 07-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Heslin, James M.  
 ; REGISTRATION NUMBER: 29,541  
 ; REFERENCE/DOCKET NUMBER: 15270-002210US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-326-2400  
 ; TELEFAX: 415-326-2422  
 ; INFORMATION FOR SEQ ID NO: 17:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 30 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-08-660-531-17

Query Match 85.7%; Score 54; DB 4; Length 30;  
 Best Local Similarity 85.7%; Pred. No. 0.00044;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KTEEISEVNXVAEF 14  
 Db 17 KTEEISEVNLDAEF 30

## RESULT 3

US-08-659-984A-16  
 ; Sequence 16, Application US/08659984A  
 ; Patent No. 5942400  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Anderson, John P.  
 ; APPLICANT: Sinha, Sukanto  
 ; APPLICANT: Jacobson-Croak, Kirsten L.  
 ; TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Ctr., 8th Floor  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94111-3834  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION NUMBER: US/08/659,984A  
 ; FILING DATE: 07-JUN-1996  
 ; CLASSIFICATION: 436  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/485,152  
 ; FILING DATE: 07-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Heslin, James M.  
 ; REGISTRATION NUMBER: 29,541  
 ; REFERENCE/DOCKET NUMBER: 15270-002810US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-326-2400  
 ; TELEFAX: 415-326-2422  
 ; INFORMATION FOR SEQ ID NO: 16:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 33 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-08-659-984A-16

Query Match 85.7%; Score 54; DB 2; Length 33;  
 Best Local Similarity 85.7%; Pred. No. 0.00049;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KTEEISEVNXVAEF 14  
 Db 8 KTEEISEVNLDAEF 21

## RESULT 4

US-08-660-531-16  
 ; Sequence 16, Application US/08660531  
 ; Patent No. 6221645  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chrysler, Susanna M.S.  
 ; APPLICANT: Sinha, Sukanto  
 ; APPLICANT: Keim, Pamela S.  
 ; APPLICANT: Anderson, John P.  
 ; TITLE OF INVENTION: Beta-Secretase  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Townsend and Townsend and Crew LLP  
 ; STREET: Two Embarcadero Ctr., 8th Floor  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94111-3834  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION NUMBER: US/08/660,531  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/480,498  
 ; FILING DATE: 07-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:

NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-15

Query Match 85.7%; Score 54; DB 4; Length 33;  
Best Local Similarity 85.7%; Pred. No. 0.00049;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
Db 8 KTEISEVNLDAEF 21

RESULT 5  
US-08-659-984A-15  
Sequence 15, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-659-984A-15

Query Match 85.7%; Score 54; DB 2; Length 42;  
Best Local Similarity 85.7%; Pred. No. 0.00065;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
Db 8 KTEISEVNLDAEF 21

RESULT 6  
US-08-660-531-15  
Sequence 15, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-15

Query Match 85.7%; Score 54; DB 4; Length 42;  
Best Local Similarity 85.7%; Pred. No. 0.00065;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
Db 17 KTEISEVNLDAEF 30

RESULT 7  
US-08-659-984A-21  
Sequence 21, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-659-984A-15

Query Match 85.7%; Score 54; DB 2; Length 42;  
Best Local Similarity 85.7%; Pred. No. 0.00065;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 506 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-659-984A-21

Query Match 85.7%; Score 54; DB 2; Length 506;  
Best Local Similarity 85.7%; Pred. No. 0.011;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
||||||| |||  
DB 398 KTEEISEVNLDAEF 411

RESULT 8  
US-08-660-531-21  
Sequence 21, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:

NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 506 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-660-531-21

Query Match 85.7%; Score 54; DB 4; Length 506;  
Best Local Similarity 85.7%; Pred. No. 0.011;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
||||||| |||  
DB 398 KTEEISEVNLDAEF 411

RESULT 9  
US-09-054-334-4  
Sequence 4, Application US/09054334  
Patent No. 6329163  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Jacobson-Croak, Kirsten L.  
APPLICANT: Sinha, Sukanto  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/054,334  
FILING DATE: 02-APR-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 015270-002820US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 506 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-054-334-4

Query Match 85.7%; Score 54; DB 4; Length 506;  
Best Local Similarity 85.7%; Pred. No. 0.011;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEEISEVNXVAEF 14



Db 398 KTEEISEVNLDAEF 411  
||||||| |||

RESULT 10  
US-09-548-372D-12  
; Sequence 12, Application US/09548372D  
; Patent No. 6420534  
; GENERAL INFORMATION:  
; APPLICANT: GURNEY ET AL.  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: 29915/62801  
; CURRENT APPLICATION NUMBER: US/09/548,372D  
; CURRENT FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: US 60/155,493  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: US 09/404,133  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: PCT/US99/20881  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: US 60/101,594  
; PRIOR FILING DATE: 1998-09-24  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 12  
; LENGTH: 695  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-548-372D-12

Query Match 85.7%; Score 54; DB 4; Length 695;  
Best Local Similarity 85.7%; Pred. No. 0.016; 2; Indels 0; Gaps 0;  
Matches 12; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
||||||| |||  
Db 587 KTEEISEVNLDAEF 600

RESULT 11  
US-09-548-367D-12  
; Sequence 12, Application US/09548367D  
; Patent No. 6440698  
; GENERAL INFORMATION:  
; APPLICANT: GURNEY ET AL.  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: 29915/6280H  
; CURRENT APPLICATION NUMBER: US/09/548,367D  
; CURRENT FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: US 60/155,493  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: US 09/404,133  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: PCT/US99/20881  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: US 60/101,594  
; PRIOR FILING DATE: 1998-09-24  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 12  
; LENGTH: 695  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-548-367D-12

Query Match 85.7%; Score 54; DB 4; Length 695;  
Best Local Similarity 85.7%; Pred. No. 0.016; 2; Indels 0; Gaps 0;  
Matches 12; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
||||||| |||

Db 587 KTEEISEVNLDAEF 600

RESULT 12  
US-09-548-372D-18  
; Sequence 18, Application US/09548372D  
; Patent No. 6420534  
; GENERAL INFORMATION:  
; APPLICANT: GURNEY ET AL.  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USE  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: 29915/62801  
; CURRENT APPLICATION NUMBER: US/09/548,372D  
; CURRENT FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: US 60/155,493  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: US 09/404,133  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: PCT/US99/20881  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: US 60/101,594  
; PRIOR FILING DATE: 1998-09-24  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 18  
; LENGTH: 697  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-548-372D-18

Query Match 85.7%; Score 54; DB 4; Length 697;  
Best Local Similarity 85.7%; Pred. No. 0.016; 2; Indels 0; Gaps 0;  
Matches 12; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
||||||| |||  
Db 587 KTEEISEVNLDAEF 600

RESULT 13  
US-09-548-367D-18  
; Sequence 18, Application US/09548367D  
; Patent No. 6440698  
; GENERAL INFORMATION:  
; APPLICANT: GURNEY ET AL.  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USE  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: 29915/6280H  
; CURRENT APPLICATION NUMBER: US/09/548,367D  
; CURRENT FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: US 60/155,493  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: US 09/404,133  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: PCT/US99/20881  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: US 60/101,594  
; PRIOR FILING DATE: 1998-09-24  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 18  
; LENGTH: 697  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-548-367D-18

Query Match 85.7%; Score 54; DB 4; Length 697;  
Best Local Similarity 85.7%; Pred. No. 0.016; 2; Indels 0; Gaps 0;  
Matches 12; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY 1 KTEEISEVNXVAEF 14  
||||||| |||  
Db 587 KTEEISEVNLDAEF 600

```

RESULT 14
US-08-371-930-25
; Sequence 25, Application US/08371930
; Patent No. 5578451
; GENERAL INFORMATION:
; APPLICANT: Nishimoto, Ikuro
; TITLE OF INVENTION: ALZHEIMER'S DISEASE THERAPEUTICS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/019,208
; FILING DATE: February 18, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/154001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-371-930-25

```

```

Query Match 76.2%; Score 48; DB 1; Length 58;
Best Local Similarity 78.6%; Pred. No. 0.014;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 KTEEISEVNXVAEF 14
   |||||
DB 37 KTEEISEVKMDAEF 50

```

```

RESULT 15
PCT-US94-01712-25
; Sequence 25, Application PC/TUS9401712
; GENERAL INFORMATION:
; APPLICANT: Nishimoto, Ikuro
; TITLE OF INVENTION: ALZHEIMER'S DISEASE THERAPEUTICS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01712
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/019,208
; FILING DATE: February 18, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/154001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
PCT-US94-01712-25

```

```

Query Match 76.2%; Score 48; DB 5; Length 58;
Best Local Similarity 78.6%; Pred. No. 0.014;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 KTEEISEVNXVAEF 14
   |||||
DB 37 KTEEISEVKMDAEF 50

```

```

Search completed: April 23, 2003, 12:46:27
Job time : 23.6897 secs

```